

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Searcher: _____

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 5/8/01

Date Completed: 5/8/01

Searcher Prep & Review Time: _____

Clerical Prep Time: _____

Online Time: _____

Type of Search

NA Sequence (#) _____

AA Sequence (#) 1

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbit _____

Dr.Link _____

Lexis/Nexis _____

Sequence Systems 02

WWW/Internet _____

Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:54:40 ; Search time 12.28 Seconds

(without alignments)
769,686 Million cell updates/sec

Title: US-09-323-597b-2

Perfect score: 2717

Sequence: 1 MALNSGSPPAIGPYENHGX.....VYGVNVTWTDWYRQNRADG 492

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCFUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	US-09-342-749-2	Sequence 2, Appl 1
2	1540	56.7	283	US-08-807-151-1	Sequence 1, Appl 1
3	676.5	24.9	435	US-09-008-271A-6	Sequence 6, Appl 1
4	660	24.3	798	US-08-200-900A-2	Sequence 2, Appl 1
5	558.5	20.6	855	PCT-US94-00616-2	Sequence 2, Appl 1
6	556	20.5	638	US-09-027-337-2	Sequence 2, Appl 1
7	539.5	19.9	416	US-08-681-151-3	Sequence 3, Appl 1
8	334	19.7	356	US-09-000-846-2	Sequence 2, Appl 1
9	528.5	19.5	256	US-08-681-151-1	Sequence 3, Appl 1
10	527.5	19.4	418	US-08-807-151-3	Sequence 2, Appl 1
11	521.5	19.2	235	US-08-807-151-3	Sequence 25, Appl 1
12	518.5	19.1	812	US-08-248-629A-1	Sequence 3, Appl 1
13	518.5	19.1	812	US-08-451-932-1	Sequence 1, Appl 1
14	518.5	19.1	812	US-08-452-260-1	Sequence 1, Appl 1
15	518.5	19.1	812	US-08-326-785-1	Sequence 1, Appl 1
16	518.5	19.1	812	US-08-612-788-1	Sequence 1, Appl 1
17	518.5	19.1	812	US-08-605-598B-1	Sequence 1, Appl 1
18	518.5	19.1	812	US-08-429-743-1	Sequence 1, Appl 1
19	518.5	19.1	812	US-08-866-735-1	Sequence 1, Appl 1
20	518.5	19.1	812	US-09-066-028-1	Sequence 1, Appl 1
21	518.5	19.1	812	PCT-US95-05107-1	Sequence 1, Appl 1
22	513.5	18.9	232	US-08-508-448C-19	Sequence 19, Appl 1
23	459.5	16.9	314	US-09-008-271A-3	Sequence 3, Appl 1
24	452	16.6	546	5200340-6	Patent No. 5200340
25	450	16.5	270	US-08-978-404B-8	Sequence 8, Appl 1
26	449	16.5	653	US-08-148-910-12	Sequence 12, Appl 1

28	449	16.5	655	1	US-08-448-937A-12	Sequence 12, Appl 1
29	447	16.5	276	2	US-09-016-366A-15	Sequence 15, Appl 1
30	447	16.5	276	2	US-08-978-404B-21	Sequence 21, Appl 1
31	446	16.4	230	1	US-08-379-621-2	Sequence 2, Appl 1
32	446	16.4	230	1	US-08-147-000B-2	Sequence 2, Appl 1
33	446	16.4	230	2	US-08-889-078-2	Sequence 2, Appl 1
34	445.5	16.4	791	1	US-08-643-219-1	Sequence 1, Appl 1
35	445.5	16.4	791	2	US-09-131-995-1	Sequence 1, Appl 1
36	445.5	16.4	791	2	US-08-823-087B-1	Sequence 1, Appl 1
37	445.5	16.4	791	3	US-08-831-350-1	Sequence 1, Appl 1
38	445.5	16.4	810	1	US-07-854-603-2	Sequence 1, Appl 1
39	445.5	16.4	810	1	US-08-147-000B-29	Sequence 29, Appl 1
40	445.5	16.4	810	6	5200340-8	Patent No. 5200340
41	444.5	16.4	790	1	US-08-469-486-54	Sequence 54, Appl 1
42	444.5	16.4	790	2	US-08-469-486-54	Sequence 54, Appl 1
43	442.5	16.3	222	1	US-08-456-840-46	Sequence 46, Appl 1
44	442.5	16.3	222	1	US-08-266-407A-46	Sequence 46, Appl 1
45	442.5	16.3	222	2	US-08-892-544-46	Sequence 46, Appl 1

ALIGNMENTS

RESULT 1
US-09-342-749-2
Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TPBRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 100.0%; Score 2717; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 2, 2e-241;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MANSOSPPIATPPYENHGXQOPENPYPAQPTVYEVHDAQYPSPPQYAPRVLTQA	60
DB	1	MALNSGSPAIPIPPYENHGXQOPENPYPAQPTVYEVHDAQYPSPPQYAPRVLTQA	60
QY	61	SNPVCTOPKPSSTVCTSKTKKALCTITLTGFLVGAALAGLMTFKGSKNSGTEC	120
DB	61	SNPVCTOPKPSSTVCTSKTKKALCTITLTGFLVGAALAGLMTFKGSKNSGTEC	120
QY	121	DSSGTCINPNCMDGVSHCPGDEENRCVRLYGPNFLLOYVSSQKSMHVCODDMMNT	180
DB	121	DSSGTCINPNCMDGVSHCPGDEENRCVRLYGPNFLLOYVSSQKSMHVCODDMMNT	180
QY	181	GRACRDMGKRNKNNYSOGIYDSDGSTSEFKLNTSAGNDVIYKKLHSDACSSKAVVSLR	240
DB	181	GRACRDMGKRNKNNYSOGIYDSDGSTSEFKLNTSAGNDVIYKKLHSDACSSKAVVSLR	240
QY	241	CIACGVNLSSROSRITYGESALPGAMPQVSLHYOVHVGSGITIPKIVTAHCVK	300
DB	241	CIACGVNLSSROSRITYGESALPGAMPQVSLHYOVHVGSGITIPKIVTAHCVK	300
QY	301	PLNPMHTAFAGILROSFMYGAGYOVERVISHPYDSKTKNNDTALMKQPLTFNDL	360
DB	301	PLNPMHTAFAGILROSFMYGAGYOVERVISHPYDSKTKNNDTALMKQPLTFNDL	360

Db 301 PLNPNHMTAFAGILRQSPMFYAGAYEVESHRYDSKTKNDIALMKLQKPLTFNDL 360
Qy 361 VPRVCLPNEGMLQPEQLCWSIGWGATEEKGTSYVLNAKVLLETRCNSRYVDNL 420
Db 361 VPRVCLPNEGMLQPEQLCWSIGWGATEEKGTSYVLNAKVLLETRCNSRYVDNL 420
Qy 421 TRAMICAGLQGVSCGDSGGLPVTSTNNIWMILGDTSMWGCCKAKRPGYGVNMF 480
Db 421 TRAMICAGLQGVSCGDSGGLPVTSTNNIWMILGDTSMWGCCKAKRPGYGVNMF 480
Qy 481 TDWYRQMRADG 492
Db 481 TDWYRQMRADG 492
RESULT 2
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 556016
; US-08-807-151-1
Query Match 56.7%; Score 1540; DB 3; Length 283;
Best Local Similarity 99.6%; Pred. No. 1.3e-133;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 61 QVSLHGVNHYVGGSLITPEWITVAHCEKPLNNPMHTAFAGILRQSPMFYAGAYE 120
Qy 330 KYISHENYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNEGMLQPEQLCWSIGATEE 389
Db 121 KYISHENYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNEGMLQPEQLCWSIGATEE 180
Qy 390 KGKTSVLAARVLLLETRCNSRYVDNLITPAMICAGLQGVSCGDSGGLPVTST 449
Db 181 KGKTSVLAARVLLLETRCNSRYVDNLITPAMICAGLQGVSCGDSGGLPVTST 240
Qy 450 NNIMWILGDTSMWGCCKAKARPGYGVNMFYTDWYRQMRADG 492
Db 241 NNIMWILGDTSMWGCCKAKARPGYGVNMFYTDWYRQMRADG 283
RESULT 3
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Tang, Tom Y.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheila
; REGISTRATION NUMBER: 41,701
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLANOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-008-271A-6
Query Match 24.9%; Score 676.5; DB 4; Length 435;
Best Local Similarity 39.1%; Pred. No. 4.8e-54;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

```

Db 75 CDGELDPLGDEDEHCVKSFPEGPAAVAVRLSKDSTLQVLDATGCMWESACDFNTEALA 134
QY 182 RAACRDQGY--KNNFYSQ-----GIYDSDGSTSPFKLNTSAGNVDIYKLYHSDACS 232
Db 135 ETACRQMGYSKPFRAVEIGPDLDLVEITENSQELRANSSG-----PCL 182
QY 233 SKAVYSLRJCAGVNLNSSRSQRIYVGSALPGAMPQVSLHYOVNHYVCGSIITPENIV 292
Db 183 SGLSVSLHCLACGSL--KTPRYVGGESAVSDMPQVSLQYDKOHVCGSILDPHMYL 239
QY 293 TAACVCEKPLNPNHMTAFAGILR-QSFMEYAGAYQVEKVI---SHPIYDSTKTKNDAL 348
Db 240 TAACHCFRKH-TDVNRMKVRAGSDKLGSE---PSLAVALKIIIEFNPMY---PKNDIAL 291
QY 349 MKLQKPLTFENDLVKPVCLPNPGMMLQPEQLCMISGMATEEK-GKTSVNLNAKLLIET 407
Db 292 MKLQPLTFSTVPIKICLFFDEELITPATPLITMGTKONGMSILLOASQVIDS 351
QY 408 QRCNSRYVDNLITPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMWLIIGDTSMGSCAK 467
Db 352 TRCNADAYQGEVEYERKMKACGIPGEGVDTGCGDGGPLMYQSDQ-WHYVGIYSMGYCGG 410
QY 468 AYRPGYGNVAFWFDWIYKRMAD 491
Db 411 PSTPGYVYTKVSAVILNWIYNVWKAE 434

RESULT 4
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Melner, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-200-900A-2

Query Match 24.3% Score 660; DB 1: Length 798;
Best Local Similarity 33.9% Pred. No. 3.8e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTQPSGPGVCTSKTKALCITLTL-----GTFLVGAALAGLMLKFMGSKCSNSG 117

```

```

Db 374 VYTGPGPVNDVEFTTNRMVLEITDNMLAKGFKANFTTGLG-----IPEPCEDN 426
QY 118 IECDSGTCINPSNMCDGVSHCPGEGEDENRCVRLYG-----PNFLLQYSSQKRSWH 169
Db 427 FQC-KDGEICILVNLCDGFPKCKDSDKAHCVRLEFNGTTDSSGLVQFRIO-----SIWH 479
QY 170 PVEDDMNENTGRAACRDQMGYKNNFYSQGIYDSDGSTSPFKLNTSAGNVDIYKLYHSD 229
Db 480 VACAENMTTQISDVCQLLGLGTG--NSSVPTFSTGGGPYVNLNTPANGSLI---LTPSQ 534
QY 220 ACSSKAVYSLK--TAGVNLNLSNOS-RIVGGSALPGAMPQVSLHYOVNHYVCGSII 286
Db 535 QLEDLSLLQCNKSKCKKLVTOEVSPIKIVGSDSREGAMPVVALFEDQVCGASLV 594
QY 287 TPENITVAHCVEKPLNPNHMTAFAGILRQSFMEYAGAYQVE-----KVISHPYDSTK 341
Db 535 SRDWLVSAHCYGGNMPSPKRAVLGLHMASNL---TSPQIETRLIQIYINPHYKRR 651
QY 342 KNNDIALKLOKPLTFENDLVKPVCLPNPGMMLQPEQLCMISGMATEEKGTSEVLNAK 401
Db 652 KNNDIAMHLEKMYVTDYIDPICLPEENQVFPGRICISAGMGALIVQGSTADVLQEAD 711
QY 402 VLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMWLIIGDTSM 461
Db 712 VPLLSNEKCOQOQMPYXN-ITENMVCAGYEAGVDSGCGDGGPLMCOENRMMLAGVTSF 770
QY 462 GSGCAKAYRPGYGNVAFWFDWI 484
Db 771 GYOCALPNRPGYARVAPRTEWT 793

RESULT 5
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-00616-2

Query Match 24.3% Score 660; DB 5: Length 798;
Best Local Similarity 33.9% Pred. No. 3.8e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTQPSGPGVCTSKTKALCITLTL-----GTFLVGAALAGLMLKFMGSKCSNSG 117
Db 374 VYTGPGPVNDVEFTTNRMVLEITDNMLAKGFKANFTTGLG-----IPEPCEDN 426
QY 118 IECDSGTCINPSNMCDGVSHCPGEGEDENRCVRLYG-----PNFLLQYSSQKRSWH 169
Db 427 FQC-KDGEICILVNLCDGFPKCKDSDKAHCVRLEFNGTTDSSGLVQFRIO-----SIWH 479
QY 170 PVEDDMNENTGRAACRDQMGYKNNFYSQGIYDSDGSTSPFKLNTSAGNVDIYKLYHSD 229
Db 480 VACAENMTTQISDVCQLLGLGTG--NSSVPTFSTGGGPYVNLNTPANGSLI---LTPSQ 534
QY 220 ACSSKAVYSLK--TAGVNLNLSNOS-RIVGGSALPGAMPQVSLHYOVNHYVCGSII 286
Db 535 QLEDLSLLQCNKSKCKKLVTOEVSPIKIVGSDSREGAMPVVALFEDQVCGASLV 594
QY 287 TPENITVAHCVEKPLNPNHMTAFAGILRQSFMEYAGAYQVE-----KVISHPYDSTK 341
Db 535 SRDWLVSAHCYGGNMPSPKRAVLGLHMASNL---TSPQIETRLIQIYINPHYKRR 651
QY 342 KNNDIALKLOKPLTFENDLVKPVCLPNPGMMLQPEQLCMISGMATEEKGTSEVLNAK 401
Db 652 KNNDIAMHLEKMYVTDYIDPICLPEENQVFPGRICISAGMGALIVQGSTADVLQEAD 711
QY 402 VLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMWLIIGDTSM 461
Db 712 VPLLSNEKCOQOQMPYXN-ITENMVCAGYEAGVDSGCGDGGPLMCOENRMMLAGVTSF 770
QY 462 GSGCAKAYRPGYGNVAFWFDWI 484
Db 771 GYOCALPNRPGYARVAPRTEWT 793

```


QY 351 LQRLPTEKNDLVKPCVLPDPCGMLQPEOLCNIISGKATEEKGKTSSEVLNAKVLLETRC 410
DB 489 LQPLPTEKNDLVKPCVLPDPCGMLQPEOLCNIISGKATEEKGKTSSEVLNAKVLLETRC 548
QY 411 NSRVYDNLTPAMICAGFLQGNVDSGSGGPLYTSKNNIMWLGDTSGSCARAYR 470
DB 549 QKRT-RBYVITKQKICAGYEGEGIDACKDSGGPLVCKHSGRMOLVITTSWEGSCAKREQ 607
QY 471 PGVYGNVMTFTDIYRQMR 490
DB 608 PGVYTKVAEYIDWLEKIQS 627

RESULT 8
US-09-000-846-2
Sequence 2, Application US/09000846
Patent No. 5981830

GENERAL INFORMATION:
APPLICANT: WU, QINGYU
APPLICANT: SADLER, JASPER
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
STREET: 2200 CLARENDON BLVD. SUITE 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: US
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,846
CLASSIFICATION: 800
FILING DATE: 30-DEC-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/866,058
FILING DATE: 30-MAY-1997

ATTORNEY/AGENT INFORMATION:
NAME: LEBOVITZ, RICHARD M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: BERLX 65P1
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 416 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-000-846-2

Query Match 19.98; Score 539.5; DB 2; Length 416;
Best Local Similarity 30.68; Pred. No. 1.7e-41;
Matches 132; Conservative 64; Mismatches 179; Indels 57; Gaps 9;

QY 77 CTSKTKKALCITLTGLFELVAGALAGLWKKFKSGSKNSGIEDSSGTCINSMWCDGV 136
DB 11 CCSRPRVVALIVTGL-LFLTG-----IGASNAIVT 40
QY 137 SHCPGGEDENCVRLYGNPILQVYSSQKSMHPVCCDDNNENYGRAACRDMGYKNFYS 196
DB 41 ILQSQDEPLVQVLSGDSRLAVFDTBSTWRLCSSRSNARVAGCGEMGLRALAH 100
QY 197 SGCIVDDSGSTSMKNTSA-----GNVDIYKKLIHSDA-----CSSAAVYSLRCIACGVN 247
DB 101 SELDVARTAGAN-----GTSGFVCEVDEGLRLAORLIDVIVSCDPPRGRLTATCCDGG-- 153

QY 248 LNSRSRIVIGESALPGAMPQVSLHVQNVHCGSITPEVITVAACVEKPLNNPMH 307
DB 154 RRLPLVRIIVIGQDSSLGRRPMQVSLRYDGTSLCGSLSGDWLTPAARCFPERNNLSR 213
QY 308 WTAFAGLRISFEMFYAGYQYEVKYSHPN-----DSKKNNDIALMKQKRLTENDLV 361
DB 214 WRFVAGAVARTSP-HAVQVQAVIYHGVLPPRPDPTIDENSNDIALVHLSSSLPLEYI 272
QY 362 KPVCLPNDGMLQPEOLCNIISGKATEEKGKTSSEVLNAKVLLETRCNSRYVDNLIT 421
DB 273 QPVCLPAAQALVDGKCYVTGNGNTQFYQQAQVLDQEARVPIISNVCNSPDEYGNQIK 332
QY 422 PAMICAGFLQGNVDSGSGGPLYV---TSKNNIMWLGDTSGSCARAYRPGVYGNV 477
DB 333 PKMECAGYPEGIDACQDGGGPFVCEDSISGTSRMLCIGIVSMGTSCALARRPGVYTRV 392
QY 478 WFTDIYRQMR 489
DB 393 TDFREWIFKAIR 404

RESULT 9
US-08-681-151-1
Sequence 1, Application US/08681151
Patent No. 5869637

GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Braxton, Scott Michael
APPLICANT: Goll, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151
FILING DATE: Herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0074US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: HEARNOT01
CLONE: 307474
US-08-681-151-1

RA Oberst M.D., Dickson R.B., Lin C.Y.;
 RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RP SEQUENCE FROM N.A.
 RA Takenuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 DR EMBL: AF118224; AAD42765.2; -
 DR EMBL: AF133086; AAF00109.1; -
 DR HSP: P00763; IDPO.
 DR MEROPS: S01.302; -
 DR INTERPRO: IPR000859; -
 DR INTERPRO: IPR001254; -
 DR INTERPRO: IPR001314; -
 DR INTERPRO: IPR002172; -
 DR PFAM: PF00057; Idl_recept_a; 4.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF00431; CUB; 2.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01209; LDLR_1; 2.
 DR PROSITE: PS00068; LDLR_2; 4.
 KM Glycoprotein; Protease.
 SQ SEQUENCE 855 AA; 94769 MW; 2614132C01F99C9 CRC64;

Query Match 20.6%; Score 558.5; DB 4; Length 855;
 Best Local Similarity 33.9%; Pred. No. 3.6e-42;
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

OY 110 GSKNSGIECDSGCTCNPSMCDGVSHCGEGDENRCVRLYGNFLTYSSQKSMH 169
 DB 522 GCSPAPATFRCSNGKLSKSGQCGKDDCGDSDGSDASCPVY---NVY----- 565
 OY 170 PQCDDMNENYGRACRDMGKNNFYSSQGIYDDSGSYFKNLTSAGNVDIYKLYHSD 229
 DB 566 -----TCT-----KHTYRCLNG-----LCLSKGNPDCDGEKDCSD 595
 OY 230 ACSSKAYVSLACICGVNLNS-SQSRIVGESALPGAMPQVSLH-VQNVHVGGSIT 287
 DB 596 GSDER-----CCDGGKRSFTROARVGGTADDEGMPQVSLHAGGCHIGCASLIS 647
 OY 288 PEMVTAHACVEKPE---LNNPMHTAFAGILRQSFMYGAGYQ---VEKYISHPNYDSK 340
 DB 648 PNLVLSAHCHTIDRGRFFYSPTQMTAFGLHDOS-QRSAPGVDERRLKRLIHPFFWDF 706
 OY 341 TKNDIALMKLQKPLTFNDLVKPYCLPMPGMLOPEOLCMTSGNGATEEKGTSEVLNAA 400
 DB 707 TFDYDIALLEKEKPAEYSSVWRPICLPDASHVFPAGKAIWVTGNGHTYGTGALLDKG 766
 OY 401 KVLLEIQRCKSRVYDNLITPAMICAGFLGAGNDSCGSDGSGGLP---VTSNNNIMWLG 458
 DB 767 EIRVINTOTCEN--LLPQOITPRMVCVGLSGVDSGCGSLSSVEADGRI-FQAGV 823
 OY 459 TSMGCGAKAYRPGVYGVVWVFTMD 484
 DB 824 VSMGDGCAQRKPKPVYTRLPFRDHI 849

RESULT 11
 OYVANG PRELIMINARY; PRT; 1379 AA.
 AC OYVANG;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CG2105 PROTEIN.
 GN CG2105.
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota: Metazoa: Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Beeson K.V., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Bertozzi D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Doolin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireit A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03840; AAF59230.1; -
 DR HSP: P00763; IDPO.
 DR FLTBASE: FBgn0033192; CG2105.
 DR INTERPRO: IPR000024; -
 DR INTERPRO: IPR000217; -
 DR INTERPRO: IPR001190; -
 DR INTERPRO: IPR001254; -
 DR INTERPRO: IPR001314; -
 DR INTERPRO: IPR002172; -
 DR PFAM: PF00057; Idl_recept_a; 2.
 DR PFAM: PF00089; trypsin; 1.
 DR PFAM: PF01392; Fz; 1.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS00227; TUBULIN; 1.
 DR PROSITE: PS01209; LDLR_1; 1.
 DR PROSITE: PS00068; LDLR_2; 2.
 SQ SEQUENCE 1379 AA; 149489 MW; A593A9CC2167E4B7 CRC64;

Query Match 20.0%; Score 544; DB 5; Length 1379;
 Best Local Similarity 33.4%; Pred. No. 1.4e-40;
 Matches 141; Conservative 65; Mismatches 158; Indels 58; Gaps 20;

OY 111 SKGNSG---IECDSGCTCNPSMCDGVSHCGEGDENRCVRLYGNF---FILDVYS 162

```
RESULT 12
US-08-807-151-3
; Sequence 3, Application US/08807151
; Patent No. 604303
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PR-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-855-0555
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 416132
; US-08-807-151-3

Query Match          19.2%, Score 521.5; DB 3; Length 235;
Best Local Similarity 42.3%, Pred. No. 3.4e-40;
Matches 99; Conservative 44; Mismatches 82; Indels 9; Gaps 3;

QY 256 IVGESALPGAMPQVSLHVNQVHVGSGSIITPEWITVAHCVEKPLNNPMHTAFAGITL 315
DB 1 IVGGDSREGAMPWVVALYFDQVCGASIVSRDVLVSAHCVGRNNEPSSKRAVLGLH 60
QY 316 ROSFMYFAGQVE-----KYISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPANG 370
DB 61 MASNN--TSPOIELRLIDQIVINPHNKRNRKNDIAMHLEMKVNYDYDQIPCLPEEN 117
QY 371 MMLPEQLCWISMGCKTESEKTSVLAANAVALLETORCNSRYVDNLITPAMICAGFL 430
DB 118 QVFPGRICISLAGKALLYOGSTADYLOEADYVPLLSNEKCOQOQMEYN-ITENMVCAGE 176
QY 431 QGNVDSGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNVFTDWI 484
DB 177 AGVDSCGDSGGPLMCOENNRMLLAGVTSFGYQCALPDRPGVYARVRFTEMI 230

RESULT 13
US-08-248-629A-1
; Sequence 1, Application US/08248629A
```

```
; Patent No. 5639725
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Angiostatin and Method of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askev
; STREET: 191 Peachtree Street, 37th floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,629A
; FILING DATE: 04/26/94
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larry W. Sculfs, Ph.D.
; REGISTRATION NUMBER: 34,025
; REFERENCE/DOCKET NUMBER: 05213-0120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 812
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-248-629A-1

Query Match          19.1%, Score 518.5; DB 1; Length 812;
Best Local Similarity 32.7%, Pred. No. 3.8e-39;
Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

QY 106 WKFMSK-CSNSGIECDSSGTCI-----NPSNMCDSVSHC--PGEDENRCVRLYGP 155
DB 446 WEYCNLKRCSERG-----GSVELPTVSQEPSSGSDSETQMGNGRDYGRKAVTAAG 499
QY 156 FLLQVYSSQSRKSMHPVC-----ODMMENTGRAACRDG---YKNFYSSGGIVDDSG 205
DB 500 TPCQGMMAOEPHRSIIFPQTPNRADLEKNYCRNPDGDVNGPMCTTN-----547
QY 206 STSPFKLNTSAGNDIYKRLHSDACSSKAVSLRCLACGVNLSNR--SRIYGESAL 263
DB 548 -----PKRLY--DYCDIPLCASASSFEQGRPOVEPKKCPGRVYGGCVAN 589
QY 264 PGAMPQVSLHVN--NVHVGSGSIITPEWITVAHCVEKPLNNPMHTAFAGILROSEMF 321
DB 550 PHSWPMQISLKRTPGQFPCGGLTAPKEMVLAHGLEKS--SRPEFYVITGAHE-----644
QY 322 YGAGYQVE-----KYISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPANGMLOPE 376
DB 645 YIRGLDVOEISVALKILPNN-----NRDIALKLRSRATITDKVIYIPACLPSPVNVADR 698
QY 377 QLCWISGKATE---EKRTSEVLAANAVALLETORCNSRYVDNLITPAMICAGFLQGN 433
DB 699 TICYTGTGETQGTFGAQLKE---AQLPVIENKVCNRVEYLNRRVSTELCAGQLAGG 754
QY 434 VDSGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNVFTDWIYRMR 489
DB 755 VDSGDSGGPLVTCFERKDYILIGVTSWGLCARPNKRGVYVRYSRFWDIERMR 810
```



```

Db 145 -PVCPLAAGTFLTGTCWYTGKATGDERDMS-VLQELAVPLDSECEKMYHTQSSSL 202
QY 415 VYDNLTPAMICAGFLOGNDVSCGDSGPLVSKNIMWILIGTSMGSCAKAYRPGVY 474
Db 203 SGERIIOGDMACAGHIDSCGDSGPLVCSINSMTGVITSMWICGARPRPGVY 262
QY 475 GNVAVFTDMYR 486
Db 263 TRVPTVDMYR 274

RESULT 14
O60235 PRELIMINARY: PRT: 418 AA.
AC O60235;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE AIRWAY TRYPSIN-LIKE PROTEASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE-98234382; Pubmed-9565616;
RA Yamaoka K., Masuda K., Ogawa H., Takagi K., Umamoto N., Yasuoka S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
RT like protease."
RL J. Biol. Chem. 273:11895-11901(1998).
DR EMBL: AB002134; BAA28691.1; -.
DR HSSP; P00730; 1P7F.
DR MEROPS; S01.301; -.
DR INTERPRO; IPR000082; -.
DR INTERPRO; IPR001254; -.
DR INTERPRO; IPR001314; -.
DR PRAM; PR00089; trypsin; 1.
DR PFAM; PF01390; SEA; 1.
DR PROSITE; PS00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PROTEASE.
KW SEQUENCE 418 AA; 46263 MW; F4BCIDB020CFBBD0 CRC64;

Query Match 19.48; Score 527.5; DB 4; Length 418;
Best Local Similarity 39.88; Pred. No. 9,5e-40;
Matches 111; Conservative 43; Mismatches 116; Indels 9; Gaps 6;
QY 212 LNTSAGNVDIYKLYHSDACSKAVSLRACIAGV--NLNSSRQSRIVGESALPGAMPW 269
Db 143 LNNH-GNLEINPST-EITSLTDAQAAWMLINEGAGPDLITLSEORITGTEAESEGSMPW 200
QY 270 QVSLHVNVAHVGCSITTPREIVYAAHCEKPLNNPWHMTAFACILNOSFFYAGVOYE 329
Db 201 QVSLRLNNAHHCGLSLNNMWILTAACF-RSNSNPDMWATSGI---STFPLRMHVR 256
QY 330 KVISHPNVDSKTKNDIALMKLQKPLTFNDLVKPYCLPNPGMILQPEOLCISMGATGE 389
Db 257 NILHNHYKSTHENDIALVLENSVFTKDIHSCVCLPAQNIIPGSTAVYTGMAOEY 316
QY 390 KCKTSEVLNAKAVLLIETORNSRYVDNLITPAMICAGFLOGNDVSCGDSGPLV-TS 448
Db 317 AGHVPPELRGOVARIISNDVCNNAHSYNGALISGLCAGVPGQGDACQDSGPLVOED 376
QY 449 KNNIMWILIGTSMGSCAKAYRPGVYGNVAVFTDMYR 487
Db 377 SRLMFTVGIYVSGDQGLPDKPGVYTRVATYLDWIRQ 415

```

```

AC O9R0W3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PLASMINOGEN PROTEIN PRECURSOR (EC 3.4.21.7).
OS PLASMINOGEN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RA Bangerter K., Johnsen A.H., Thorsen S.;
RT "Rat plasminogen: cDNA and gene structure."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE-91250378; Pubmed-1645711;
RA Kanalas J.J., Makker S.P.;
RT "Identification of the rat Heymann nephritis autoantigen (GP330) as a
RT receptor site for plasminogen."
RL J. Biol. Chem. 266:10825-10829(1991).
DR EMBL; AJ242649; CAB46014.1; -.
DR HSSP; P00747; 1P7K.
DR INTERPRO; IPR000001; -.
DR INTERPRO; IPR001254; -.
DR INTERPRO; IPR001314; -.
DR INTERPRO; IPR001400; -.
DR INTERPRO; IPR003014; -.
DR PRAM; PF00024; PAN; 1.
DR PFAM; PF00051; kringl; 5.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PRO0018; KRINGLE.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00021; KRINGLE_1; UNKNOWN_5.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PROSITE; PS00338; SOMATOSTATIN_2; UNKNOWN_1.
DR PROSITE; PS00070; KRINGLE_2; 5.
KW Signal; Hydrolase.
FT SIGNAL 1 19
FT CHAIN 20 812
FT SEQUENCE 812 AA; 90535 MW; 8C703C51410EBC9E CRC64;

Query Match 19.28; Score 522.5; DB 11; Length 812;
Best Local Similarity 27.88; Pred. No. 6,3e-39;
Matches 169; Conservative 64; Mismatches 189; Indels 187; Gaps 23;
QY 8 PPAIGPYE-----NHGYOPENPYPAQ----- 29
Db 266 PPPPGPTQCLKNGENYRGTVSVTASGKTCORHSEGTTPRHNRTPPN-FPCNLEENVC 334
QY 30 ---PTVVPYVY-----EYHPAQQYPPVPDYARVLTQASNPVCT 67
Db 325 RNPGETAPMCTYDSQLRWYECIFPCGSSVSPDQSDSVLPDQTFVQGE-----CY 377
QY 68 Q--FKPSGIVCTSKTKKALCITLTITLTVGAALAGLAKTRMGSCS----- 114
Db 378 QGNGKSYRGTSSTNT-----GKCGQSVAMTPRSHS 409
QY 115 -----NSGLECDSSGCTINPSN-----WC-----DGVSHCPGDEENRC 148
Db 410 KTRANFPAAGLENN--YCRNPNDQDGPMPCTTDPVRYEYCNLKRCSITGGVAESA 466
QY 149 VRLXPNFTIQYVSSQKSWHPVCCDDMNENYGRACRDYKYNNNFYSSGIVDDSGSTS 208
Db 467 VP-----QVPSAPGTSSETDCMYGNGKEYRGKTAVTAGTFCQEAAG--EPHSHRIF 516
QY 209 FMKINTSAGNVDI-----KLYHSDACSSKAVYSLRACICGVNLNS 250

```

Oy	241	C I A C G V M L N S R O S R I Y G E S A L P G A P M Q V O S L H V Q W H V C G S I T P E M I Y T A A C H V E K	300
Db	240	C I E C G V R - S Y K R O S R I Y G L N A S B D M P M Q V S L H V G O V H V C G S I T P E M I Y T A A C H V E	298
Oy	301	P L A N P H M W T A F A G I L R O S F M F Y G A G I O V E K Y I S H P N T D S T K A N N D A L M K L A P L F P N D I	360
Db	299	P L S G P R W T F A G I L R O S L M F Y G R H Q V E K Y I S H P N D S K T K A N N D A L M K L O T P L F N D I	358
Oy	361	V K P C L E N P G M L O P E O L C W T S G M G A T E E K G T S E Y V A N A K V L L I E F O R C N S R Y V D N L I	420
Db	359	V K P C L E N P G M L D L D E C H I S C M G A T Y E K G K T S D V L A A M V L I E P S K C N S K Y T I N N L I	418
Oy	421	T P A N I C A G L O G N V S C O G D S G G P L V T S K N N I M W L J G D I S W S G S C A K A Y A P G V Y G N V M F	480
Db	419	T P A I C A G F L O G S V D S C O G D S G G P L V T L K N G I M W L I G D I S M S G S C A K A L P G V Y G N V F	478
Oy	481	T D W I Y R O M R A D	491
Db	479	T D W I Y Q O M R A N	489

RESULT	2		
ID	090Y82	PRELIMINARY;	PRT; 490 AA.
AC	090Y82;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	MOSAIC SERINE PROTEASE EPITHELINASIN.		
GN	TMPSR52		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
OX	NCBI_Taxid-10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C;		
RA	Jacqueline E.J., Rao N.V., Rao G.N., Hoidal J.R.;		
RT	"A novel mosaic serine protease, epitheliasin.";		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF113596; AAF1308.1; -		
DR	HSSP; P00761; IAKS.		
DR	MGI; MGI:1354381; Tmpsr52.		
DR	INTERPRO; IPR001254; -		
DR	INTERPRO; IPR001314; -		
DR	INTERPRO; IPR002172; -		
DR	PFAM; PF000057; Idol_recept_e; 1.		
DR	PFAM; PF000089; trypsin_1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.		
DR	PROSITE; PS00135; TRYPSIN_SER; 1.		
DR	PROSITE; PS01209; LDLRA_1; 1.		
DR	PROSITE; PS50068; LDLRA_2; 1.		
DR	PROTEASE.		
SO	SEQUENCE.	490 AA; 5333 MW; 700AE022EB37F69F CRC64;	

	Query Match	77.9%	Score 2116:	DB 11:	Length 490:
	Best Local Similarity	78.4%:	Pred. No. 5.5e-184:		
	Matches 385;	Conservative 40:	Mismatches 64:	Indels 2:	Gaps 2:
OY	1	MAALSGSPPAICGPPYEENHGYOPENNYPAQPTVPVPUVEHYEHPAQYYSPRPQYAPRVLTGA	60		
Dd	1	MAALSGSPPGIGPCYEENHGYOSENHCPRPRAAPRAGYNLYRQQIIPSPRYQAPARITTTGA	60		
OY	61	SNPVYCOTPKKBPSETCTSKTKKAICTTLTGTFLVGAAALAAGLIMKEGSKCSNSGIEC	120		
		:			
Dd	61	STVIHTHPKS - SGACPCKSKSKSICALALATGVTLTGAAVAAYLLMREFDSCNSTSEMEC	119		
OY	121	DSASTCINPSSMWCGSVSHCPGGEDENRCVRILYGPFIIQYVSSQKKSMHPVCODDMNEXY	180		
		:			
Dd	120	GSLSLTCJSSSLWLCGVAAICPGEDEENRCVRILYGOSEFIQYVSSQKKAMYPVCODDMSENY	179		

OY	181	GRACRDMGYNNNFYSOGIYDSDGSTSEFKMLNTSAGNVDIKKLYLHSDSCSAVAYSLR	240
Dd	180	GRACKDMGKYNNNFYSOGIDIPDGSATSEFKMLNVSOGNDLYKKLYLHSDSCSSFMAYSLR	239
OY	241	CIAQGVNLNSSROSRIYGESALPGAPMOWSLHYOVNVHCGGSIITPEWIVTAAHCVER	300
Dd	240	CIEEGVR-SVAROSRIYVGVLNASPGDMPOVSLHYOQVNHCGGSIITPEWIVTAAHCVEE	298
OY	301	PLANNPWHWTAFAGILROSEFMEYAGAYOEKVISHPNYDSKTKNNDIALMKLOKPLETENDL	360
Dd	299	PLSGPRYTAFAGILROSLIMEFYGSRHQYEKVISHPNDYDSKTKNNDIALMKLOTPLEAENDL	358
OY	361	VKPCLPBPNGMMDLPEDLCRTISGNGATEBEKGTSEVUNAAKVLLETRQCRNSRYVNOLI	420
Dd	359	VKPCLPBPNGMMDLDDEOCWTSIGNGATYEKGKTSDEVUNAAVPLEIFESCKSKSYINLLI	418
OY	421	TPAMICAGELOGANDSOCGGSDGGLYNTSKNNIMWLIGDTSMGSCGARAYBPGYGNMVFE	480
Dd	419	TPAMICAGEFLOGASVSDSCGGSDGGLVTLLKNGIMWLIGDTSMGSCGARALRPYGYGNVTFE	478
OY	481	TDMYIRQWRAD	491
Dd	479	TDMYIQCMKRAN	489

Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99
Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99
Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99
Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99
Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99
Q0	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q																																																																																						

	Query Match	Similarity	77.5%	Score 2107	DB 11	Length 490
	Best Local	Similarity	78.0%	Pred. No. 3,66-183		
	Matches	383	Conservative	43	Mismatches	63
					Indels	2
					Gaps	2
QY	1	MALNSGSPPAIGPYEYENNGYOPENDPYPAQPTVVPVYEVHPAQYPPSPVPOYAPRYLTOA	60			
Db	1	MALNSGSPPGGPGPCPEENNGYSEHICPPRPAPVAPNGVYMLPYAQPSPVPOAPRITTOA	60			
QY	61	SNPVYCTOPKPSBPGVCTSKTKKALCITLTGTFTFYGAALAGLAKMKFGSKSCSNGIEC	120			
Db	61	STSVYHTHPKS - SGAPCTSKSKKSKLCIALALGTLVTLGAANAVALLMRRFDSNCSISEMBC	119			
QY	121	DSGTCINPSMWCDDVSHCPGEGEDENRCVRLYGPNFILQVYSSQKSMHPVCCODMENEY	180			
Db	120	GSSGCTCSSSSLMTCGVAVACPNGEDENRCVRYLQGSFLLQVYSSQKRAYPVCCODMSESY	179			
QY	181	GRACRDMGYNKNFNSSGGIVDDSGSTFPMKLNTSAGNVDIYKIKLYHSDASKRAYSLR	240			
Db	180	GRACKCKMGYKNFNPSSGIGIPDQSGATISFMKLNTSSGNVDLYKIKLYHSDSCSRMYSLR	239			
QY	241	CIAACGVNLNSTRQSHFVGGESALPGAMPVQVSLHQNHYVCGGSLITTEPMYITAAHCYEK	300			
Db	240	CIEGGVR - SVKRSQGIYVGLNAPSPDWMQVSLHQNHYVCGGSLITTEPMYITAAHCVEE	298			

092319	RESULT	7	
092319			
ID	092319	PRELIMINARY:	PRT; 1113 AA.
AC	092319:		
DT	01-MAY-1999	(TREMBL)	10. Created
DT	01-MAY-1999	(TREMBL)	10. Last sequence update
DT	01-JUL-2000	(TREMBL)	15. Last annotation update
DE	LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4		
	MS musculus (Mouse).		

RESULT	8	
09JUI7		
ID	09JUI7	PRELIMINARY; PRT; 855 AA.
AC	09JUI7;	
DT	01-OCT-2000 (TREMblrel. 15, Created)	
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)	
DT	01-OCT-2000 (TREMblrel. 15, Last annotation update)	
DE	MEMBRANE BOUND SERINE PROTEASE.	
GN	MSP.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus	

Tue May 8 11:56:56 2001

us-09-323-597b-2.raii

Page 9

Db 699 TICTTTTGGEGTCTFGAGRLKE-----AQLPIENKVKYNNRYEILNNRKSTELCAGLQAGG 754

QY 434 VDSQGDGGGGLVYTSKNNIMWLIDITSGSGCAKAKAYRPGYVGNVETDIIYKQMR 489

Db 755 VDSQGDGGGGLVYCFEKDKYILQGVTSMLGCAAPNKPGRVYVRSFVDIEREMR 810

Search completed: May 8, 2001, 11:55:00
Job time: 20 sec

FT	CARBOHYD	353	353		N-LINKED (GLCNAC. . .)
FT	CARBOHYD	450	450		N-GLCNAC. . .)
FT	CARBOHYD	491	491		N-LINKED (GLCNAC. . .)
FT	ACT_SITE	431	431		
FT	ACT_SITE	480	480		
FT	ACT_SITE	575	575		
FT	DISULFID	20	103		
FT	DISULFID	29	29		
FT	DISULFID	46	76		
FT	DISULFID	50	56		
FT	DISULFID	110	193		
FT	DISULFID	136	165		
FT	DISULFID	140	146		
FT	DISULFID	200	283		
FT	DISULFID	226	235		
FT	DISULFID	230	236		
FT	DISULFID	291	374		
FT	DISULFID	317	346		
FT	DISULFID	321	327		
FT	DISULFID	339	339		
FT	DISULFID	380	500		
FT	DISULFID	416	432		
FT	DISULFID	514	581		
FT	DISULFID	545	560		
FT	DISULFID	571	599		
FT	VARIANT	301	301		
FT	CONFLICT	226	226		
SO	SEQUENCE	625 AA;	70109 MM;		

F -> L (IN F11 DEFICIENCY).
 /FTID=VAR_006622.
 C -> S (IN REF. 2).
 147AFA94B7709E8F CRC64,

INTERCHAIN. (BETWEEN HEAVY AND LIGHT CHAINS).

WITH A CYS RESIDUE.

Query Match	19.6%;	Score 533;	DB 1;	Length 625;
Best Local Similarity	30.5%;	Pred. No. 1.9e-32;		
Matches 122;	Conservative 63;	Mismatches 129;	Indels 86;	Gaps 11;

[illegible]

Search completed: May 8, 2001, 11:57:42
Job time: 132 sec

```
DR Pfam: PR00089; trypsin; 1.
DR PRINTS; PR00005; APPELDOMAIN.
DR PROSITE; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin.SER. 1.
DR PROSITE; PS00135; TRYPsin.SER. 1.
DR PROSITE; PS00495; APPEL. 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT ACT_SITE 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 343 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;
```

```
Query Match 19.8%; Score 538; DB 1; Length 638;
Best Local Similarity 37.1%; Pred. No. 8.3e-33;
Matches 115; Conservative 52; Mismatches 117; Indels 26; Gaps 6;

DR EMBL; AF030065; AAB84221.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.24; -.
DR MGD; MGI:1196620; Hpn.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR Pfam; PR00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin.SER. 1.
DR PROSITE; PS00135; TRYPsin.SER. 1.
KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 202 416 CATALYTIC.
FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 187 203 INTERCHAIN (BY SIMILARITY).
FT DISULFID 321 337 BY SIMILARITY.
FT DISULFID 348 380 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 416 AA; 44739 MW; 432194FE4004B848 CRC64;
```

```
Query Match 19.7%; Score 536.5; DB 1; Length 416;
Best Local Similarity 30.6%; Pred. No. 6.7e-33;
Matches 132; Conservative 63; Mismatches 180; Indels 57; Gaps 9;
```

```
Query Match 19.7%; Score 536.5; DB 1; Length 416;
Best Local Similarity 30.6%; Pred. No. 6.7e-33;
Matches 132; Conservative 63; Mismatches 180; Indels 57; Gaps 9;
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 8, 2001, 11:54:40 ; Search time 23.01 seconds
(without alignments)
1222.265 Million cell updates/sec

Title: US-09-323-597B-2

Perfect score: 2717
Sequence: 1 MALNSGSPPAIGPYENHGY.....VYGNVMTDTWYRQRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID56/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT:*
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID56/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID56/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2717	100.0	492 21	Y77726 Human tumour suppressor
2	2717	100.0	492 21	Y44406 Human 20P1F12-GTC2
3	2709	99.7	492 21	Y92050 HPCa6/7 polypept
4	2696	99.2	492 21	B36801 Human TMPS22 prot
5	2696	99.2	492 21	Y57280 OVR15 homolog pro
6	1540	56.7	283 21	Y81492 Human prostate-ss
7	888	32.7	454 21	B32246 Tumour associated
8	885	32.6	452 20	Y41694 Human PRO382 prote
9	884	32.5	453 21	B44250 Human PRO382 (UNQ3
10	742.5	27.3	327 21	Y72093 Human serine prote
11	676.5	24.9	435 20	Y06437 Human protease HUP

12	674	24.8	432	21	Y99417 Human PRO1570 (UNQ
13	666.5	24.5	480	21	B08950 Human secreted pro
14	665.5	24.5	445	21	B11698 Mouse serine prote
15	660	24.3	798	15	R57283 Bovine enterokinas
16	651.5	24.0	457	21	B11699 Human serine prote
17	649.5	23.9	248	21	B43572 Human cancer assoc
18	648.5	23.9	414	21	B08912 Human secreted pro
19	601	22.1	1042	21	Y44426 Human serine prote
20	588	21.6	311	21	B11696 Mouse serine prote
21	587.5	21.6	273	21	B11696 Mouse serine prote
22	577.5	20.3	1113	21	Y44427 Mouse serine prote
23	556.5	20.6	683	21	B19551 Human matrilipase (
24	558.5	20.6	762	21	Y90284 Human peptidase, H
25	558.5	20.6	855	20	Y06671 Tumour antigen der
26	558.5	20.6	855	21	B19552 Human matrilipase.
27	558	20.5	296	21	Y72108 Human serine prote
28	558	20.5	372	21	Y72092 Human serine prote
29	544	20.0	238	21	B11695 Mouse serine prote
30	539.5	19.9	416	20	Y43325 Mouse hepsin prote
31	539.5	19.9	416	20	W96812 Human PRO618 prote
32	538.5	19.8	802	20	Y41710 Human PRO618 (UNQ3
33	538.5	19.8	802	21	B44266 Human PRO618 prote
34	538.5	19.8	802	21	B24052 Human PRO618 prote
35	534	19.7	356	19	W46917 Amino acid sequenc
36	527.5	19.4	418	17	R89435 Trypsin-like enzym
37	527.5	19.4	418	20	Y29498 Human lung tumour
38	527.5	19.4	418	20	Y29501 Human lung tumour
39	527.5	19.4	418	20	Y29502 Human lung tumour
40	527.5	19.4	418	21	B44437 Human lung tumour
41	527.5	19.4	418	21	B44438 Human lung tumour
42	526.5	19.4	418	21	B44428 Human lung tumour
43	518.5	19.1	812	16	R83959 Complete mouse pla
44	518.5	19.1	812	17	W07585 Murine plasminogen
45	518.5	19.1	812	20	Y08686 Murine plasminogen

ALIGNMENTS

RESULT 1
ID Y77726 standard; Protein; 492 AA.
XX Y77726;
AC
XX
XX 12-MAY-2000 (first entry)
DT
XX
DE Human tumour suppressor TMPS22 polypeptide.
XX
XX Tumour suppressor gene; TMPS22; cancer; human; drug design;
KW gene therapy; protein therapy.
XX
XX Homo sapiens.
OS
XX
XX WO20000605-A1.
FN
XX
PD 06-JAN-2000.
PD
XX
PE 29-JUN-1999; 99WO-US14622.
PE
XX
PR 29-JUN-1998; 98US-0091044.
PR
XX
XX (MYRI-) MYRIAD GENETICS INC.
PA
XX
PI Wong AKC, Tavtigian SV, Teng DHF;
PI WPI: 2000-170914/15.
DR N-PSDB: 287786.
XX
XX Novel tumor suppressor TMPS22 used for the diagnosis and prognosis of
PT human cancer -
XX
PS Claim 55; Page 77-79; 89pp; English.

Db 348 FLRLSMGSPTR---AYGTGSSG---YSLRLCNTGDNVCTTKSTRIVGNTSSMGE 401
 QY 267 WPMQVSLHVO---NHVCGSITTPETIVTAACV- KPLNNPHMTAFAGILROSMFY 322
 Db 402 WPMQVSLHVO---NHVCGSITTPETIVTAACV- KPLNNPHMTAFAGILROSMFY 322
 QY 323 GAGY-QVEKISHPNYDSKRNNDIALMKLOKPLTFENDLVKPLCPHGMGLPEQCLWI 381
 Db 460 DTFPSQIKETIITQNKVSEGNDHDIKLOAPLNTYEFQKPCLPKSGDTSTIYTCWV 519
 QY 382 SCGWATEEKSEKTEVLAFAVLLIETORCNSRYVDNLITPAMICAGFLQGVDSOCGS 441
 Db 520 TGMGFSKEKEIGNILQKVPVLTNECCQKRY-QDKKITQRYACGKRGADACGDS 578
 QY 442 GGPLVYSKNNIMMLIGDTSMGSCAKARYPGVYGVNVTFTDIYRQNR-ADG 492
 Db 579 GGPLVCKHNGMRLVGTITSMGECARREGQVYTKVAYMDMILEKTOSSDG 630

RESULT 9
 HEP5_HUMAN STANDARD; PRT; 417 AA.
 AC P05981;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE
 1).
 DE HPN OR TMPRSS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=88209431; PubMed=2835076;
 RA Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
 RT "A novel trypsin-like serine protease (hepsin) with a putative
 RL transmembrane domain expressed by human liver and hepatoma cells.";
 RL Biochemistry 27:1067-1074(1988).
 RN (2)
 RP CHARACTERIZATION.
 RX MEDLINE=91358502; PubMed=1885621;
 RA Tsui J.A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
 RA Chou S.H., Kurachi K.;
 RT "Hepsin, a cell membrane-associated protease. Characterization,
 RL tissue distribution, and gene localization";
 RL J. Biol. Chem. 266:16948-16953(1991).
 RN (3)
 RP CHARACTERIZATION.
 RX MEDLINE=93348237; PubMed=8346233;
 RA Torres-Rosado A., O'Shea K.S., Tsui J.A., Chou S.H., Kurachi K.;
 RT "Hepsin, a putative cell-surface serine protease, is required for
 RL mammalian cell growth.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 CC OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST
 CC LEVEL IN LIVER.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL, M18930; AAA36013.1; .

DR EMBL: X07732; CAA30558.1; .
 DR EMBL: X07002; CAA30058.1; .
 DR PIR: S00845; S00845.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; .
 DR MIM: 142440; .
 DR InterPro: IPR001254; .
 DR InterPro: IPR001314; .
 DR Pfam: PF00089; trypsin. 1.
 DR PRINTS: PF00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 162 NON-CATALYTIC CHAIN (POTENTIAL).
 FT CHAIN 163 417 CATALYTIC CHAIN (POTENTIAL).
 FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 18 44 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 163 417 CATALYTIC.
 FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 188 204 BY SIMILARITY.
 FT DISULFID 322 338 BY SIMILARITY.
 FT DISULFID 349 381 BY SIMILARITY.
 FT CARBOHYD 112 112 N-LINKED (GLCNAC...) (POTENTIAL).
 SO SEQUENCE 417 AA; 45011 MW; B2086F661E51D7 CRC64.

Query Match 20.8%; Score 565.5; DB 1; Length 417;
 Best Local Similarity 30.7%; Pred. No. 4,6e-35;
 Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;

QY 77 CTSKTKKALCTITLTGTFVGAALAGLMLKFMGSKNSGICDSSGICINPNNCDGV 136
 Db 12 CCSRPRVYA--ALTFAGTLTLTAIGA-----ASMAIYA 41
 QY 137 SHCPGGEDENCVRLYGPNILOVYSSORSMHPVCOODMNNENYGRACRDMGKNNFY 196
 Db 42 VILRSQDEPLIPYQVSSADRLAVFQKTEGTWRLLCSSSNARVALSCSEMGFLALH 101
 QY 197 SGQIVDDSGSTSMKINTSAGNVDIK-KLYHSDA-----CSKAVSLRCAICGV 246
 Db 102 SELDVTAGA-----NGSMFCFVDEGRPLPHORLEVIYVCDPCRGFLAICDDCG- 154
 QY 247 NLNSSRSQRVYGESALPGAMPQVSLHONVHVCSSITTPETIVTAACVKEKPLNNW 306
 Db 155 -RRKLVDRIYVGRDTSLSGMPQVSLRYDGAHLGSSLSLSDGMYLTAAHCFERRNVLS 213
 QY 307 HMTAFAGILROSMFYGAGYQVEKISHPNY-----DSKTKNDIALMKLOKPLTFNDL 360
 Db 214 RMVAFGAVAGASP-HGLQLGQVAVYHGGYLPFRDPNEBNSNDIALVHSSPLRTY 272
 QY 361 VKPCLPNEGMLQPEQLCWISGWTAEKTSVLAARVLLIETORCNSRYVDNL 420
 Db 273 ICPVCLPAAGQALVDDKICTVGTWGTGYGOAGVLOEARVPIISNDVCGNADFGNDI 332
 QY 421 TPAMICAGFLQGVDSOCGDSGLV-----TSKNNITWMLIGDTSMGSCAKARYPGVYN 476
 Db 333 KPMFCAGYPEGIDACGDSGGGPFVCEDSISRTPRMRLGIVSMGTGALOKRPGVYTK 392
 QY 477 VAVFTDIYRQNR 489
 Db 393 VSDFRMIFQAK 405

RESULT 10
 ST14_HUMAN STANDARD; PRT; 855 AA.
 ID ST14_HUMAN
 AC Q91576;
 DT 01-OCT-2000 (rel. 40, Created)

OY 481 TDWYRQMRADG 492
 DB 481 tdwlyrqmrads 492

RESULT 3

Y92050 Y92050 standard; Protein; 492 AA.

AC Y92050;

DE 01-AUG-2000 (first entry)

DE HRPc6/7 polypeptide from androgen-inducible gene clone.

KM Androgen inducible; testosterone; prostate cancer; cytostatic;
 KM TMPRSS2; diagnosis.

OS Homo sapiens.

PN W0200018961-A2.

PD 06-APR-2000.

PF 30-SEP-1999; 99MO-US22535.

PR 30-SEP-1998; 98US-0163759.

PR 30-SEP-1998; 98US-0164159.

(MILL-) MILLENNIUM PHARM INC.

PI Macbeth KJ, Shyjan AW;

PI WPI: 2000-293182/25.

DR N-PSDB; A08803.

PT Novel methods for identifying compounds for treating prostate cancer
 PT comprising measuring the level of expression or activity of 1 or more
 PT of 11 genes or their products

PS Claim 2; Fig 3; 108pp; English.

CC This protein is encoded by a gene which is androgen (e.g. testosterone)
 CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)
 CC and constitutively expressed in androgen-independent prostate cancer
 CC cells (e.g. LNCaP cells). Agents which decrease the expression or
 CC activity of these clones may slow or arrest the growth of prostate cancer
 CC cells or may kill them. HRPc6/7 can be obtained from the sequence of
 CC the known gene for TMPRSS2. A compound useful for treating prostate
 CC cancer can be identified in a novel method comprising measuring the
 CC expression level, or activity, of HRPc6/7, 3, 6/7, 8, 9, 10, 13, 14, 15,
 CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell. In the
 CC presence and absence of a test compound. The sequences may also be used
 CC in diagnosis of prostate cancer and to determine efficacy of treatment
 CC for prostate cancer.

CC Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 21; Length 492;
 Best Local Similarity 99.8%; Pred. No. 1.1e-202;
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNNGSPRAIGPYENHGVOPENPYPAQPPVPTVEVHNAQYPPSPVYAPRVLTQA 60

DB 1 malnngspapaypyenhygpenyppaqpvptvyevehpapypvpyaprvltqa 60

OY 61 SNPVYCQPKSPSGTCTSKTKKALCTLTGLTFLVGAALAGLIMFGSKSCNSGTEC 120
 DB 61 snpvyctqpkspsgtctsktkkalcitltglftlvgaaalagllmfgskscnsgtec 120

OY 121 DSSGTCINPSNMGDGVSHCPGGEDENRCLVLYGPNFIQVYSQKRSWHPVCQDDMNENY 180

DB 121 dssgtcinpshnmcgvshcpgeedenrcvrllygpnfiqvysqkrswhpvcqddmneny 180
 OY 181 GRAACRDMGKKNFNFSSQGIYDSDGSTFMRKLTMSAGNVDIYRKLXHSNACSSKAVVSLR 240
 DB 181 graacrdmgyknfnfssqglivdsdgsfmrkltmsagnvdiylrklxhsnacsksavvslr 240
 OY 241 CIACGVNLSSRSRIRIVGSSALPGANPKOVSLHVQNVHVGSGSIITPEVITPAHCVER 300
 DB 241 ciacgvnlssrsrigrivgssalpgawpovslhvqnvhvgsgsiltpevltcaahvev 300
 OY 301 PLNNPWHHTAFACILRSGFMYGAGYOVEKVISHPNVDSTKKNNDLAKMLQPLFENDL 360
 DB 301 plnnpwhrtafacilrsgfmygagyevekvishpnvdskknndlakmlqplfctndl 360
 OY 361 VKPYCLPFGMMLQPEOLCWINSGWATEEKGKTSSEVNAARVLLIETQRONSRYVDNLI 420
 DB 361 vkpyclpfpmmllqpeqlcwlisgwateekgktsesevlnaakvllietqrnsryvndll 420
 OY 421 TPANICAGFLQGNVDSGQSGGPELVTSKNNIMWLIDTISWGSQCAKAPYGVNVNMF 480
 DB 421 tpamicagflqgnvdsqsggpevltsknnlwlldtswsgcakaypygvnmvf 480
 OY 481 TDWYRQMRADG 492
 DB 481 tdwlyrqmrads 492

RESULT 4

ID B36901 B36901 standard; Protein; 492 AA.

AC B36901;

DE 26-FEB-2001 (first entry)

DE Human TMPRSS2 protein.

KM Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;
 KM PART-1; neoplastic.

OS Homo sapiens.

PN W0200065067-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000MO-US10920.

PR 23-APR-1999; 99US-0130778.

PR 30-AUG-1999; 99US-0151585.

PR 30-DEC-1999; 99US-0174003.

PR 24-JAN-2000; 2000US-017751.

(UNIV) UNIV WASHINGTON.

PI Nelson PS, Hood L, Lin B;

PI WPI: 2000-679676/66.

DR N-PSDB; C83325.

PT Polynucleotide encoding prostate specific androgen regulated
 PT polypeptides and inhibitor of the peptides useful for treating or
 PT reducing the progression of prostate neoplastic condition in an
 PT individual -

PS Claim 63; Page 86-88; 121pp; English.

CC The present invention relates to prostate specific androgen regulated
 CC proteins. The invention may be used to determine an expression level
 CC of the prostate-specific proteins ARSDRL, TMPRSS2, or PART-1 in a
 CC fluid sample or prostate cell sample from an individual. It may also
 CC be used for diagnosing and predicting the susceptibility of a

```

RESULT 7
ENTR_PIG STANDARD: PRT: 1034 AA.
AC 01-FEB-1996 (rel. 33, Created)
BT 01-FEB-1996 (rel. 33, Last sequence update)
DT 15-JUL-1998 (rel. 36, Last annotation update)
DE ENEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS57 OR ENTR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Duodenal mucosa;
RA MEDLINE=94327548; PUBMED=8051081;
RA Matsumura M., Ichinose M., Yanagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shimomura K.,
RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;
RA "Structural characterization of porcine enteropeptidase.";
RA J. Biol. Chem. 269:19976-19982(1994).
RL - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROPEASE.
CC - PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isdb-sb.ch/announce/
CC or send an email to license@isdb-sb.ch).
CC -----
CC EMBL: D30799; BA006459.1; -
CC HSSP: P00763; IDPO.
CC -----
CC MEROPS: S01.156; -
CC InterPro: IPR000082; -
CC InterPro: IPR000859; -
CC InterPro: IPR000998; -
CC InterPro: IPR001190; -
CC InterPro: IPR001254; -
CC InterPro: IPR001314; -
CC InterPro: IPR002172; -
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF00629; MAM; 1.
CC Pfam: PF01390; SEA; 1.
CC Pfam: PF00530; SRCR; 1.
CC Pfam: PF00057; Id1_recept_a; 2.
CC Pfam: PF00089; Trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS00740; MAM_1; 1.

```

```

DR PROSITE: PS50060; MAM_2; 1.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 2.
KV Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KV Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 52 117
FT CHAIN 118 799
FT CHAIN 800 1034
FT TRANSMEM 19 47
FT DOMAIN 197 238
FT DOMAIN 240 349
FT DOMAIN 357 519
FT DOMAIN 539 649
FT DOMAIN 656 694
FT DOMAIN 693 786
FT ACT_SITE 840 840
FT ACT_SITE 891 891
FT ACT_SITE 986 986
FT LIPID 2
FT DISULFID 199 212
FT DISULFID 206 225
FT DISULFID 219 236
FT DISULFID 658 670
FT DISULFID 665 683
FT DISULFID 677 692
FT DISULFID 787 911
FT DISULFID 825 941
FT DISULFID 925 992
FT DISULFID 956 971
FT DISULFID 982 1010
FT CARBOHYD 116 116
FT CARBOHYD 147 147
FT CARBOHYD 170 170
FT CARBOHYD 194 194
FT CARBOHYD 283 283
FT CARBOHYD 333 343
FT CARBOHYD 350 350
FT CARBOHYD 403 403
FT CARBOHYD 455 455
FT CARBOHYD 485 485
FT CARBOHYD 518 518
FT CARBOHYD 549 549
FT CARBOHYD 645 645
FT CARBOHYD 697 697
FT CARBOHYD 701 701
FT CARBOHYD 721 721
FT CARBOHYD 740 740
FT CARBOHYD 761 761
FT CARBOHYD 804 804
FT CARBOHYD 863 863
FT CARBOHYD 902 902
FT CARBOHYD 964 964
SQ SEQUENCE 1034 AA; 114776 MM; 0388C64CF64CC368 CRC64;

Query Match 23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity 33.3%; Pred. No. 1e-39;
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;

OY 65 VCTQPKSPSGIVCTSKTKKALCIT---LTG---TFLVGAALAGLLMKRGSKCSNSG 117
DB 610 VYTGPEVEDVESTNRMTVLEITNDALTKGGFKANFTGTHLG-----IPERCKDN 662
OY 118 IECDSGNCINSMNCDGSHCGGDEHRCVRLG--PNFILOYSSQSRKSMHPVCOOD 175
DB 663 FQCE-NGECVLLVNLCDGSHCKDSDEHCVRLNGRANNSGLVQFPIQISIMHACKEN 721
OY 176 WNEENTGRACRDMGYKNN-----FYSSQGIYDSDGTSFMKLNISAGNWDYIKLYHSDA 230
DB 722 WTTQNSDDVCQLGLGTGNTSSMPFS-----SGGPFVRLKTPANPSLSL---LTASEQ 771
OY 231 CSKAVVSLRC--IACGVNLNSSROS-RIVGGSALPGAMPVQVSLHQNVAHVGGSITIT 287

```

```

XX 18-JUL-2000 (first entry)
XX Human prostate-associated protease (HUPAP).
DE Human prostate-associated protease; HUPAP; kallikrein; serine protease;
XX gastrointestinal disorder; cancer; prostate disorder.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 4 /note= "N-glycosylated"
XX Misc-difference 235 /label= unknown
XX /note= "Encoded by CNR"
XX
XX US6043033-A.
XX
XX 28-MAR-2000.
XX
XX 27-FEB-1997; 97US-0807151.
XX PF
XX 27-FEB-1997; 97US-0807151.
XX PR
XX (INCY-) INCYTE PHARM INC.
XX PA
XX Bandman O, Lal P;
XX PI
XX WPI; 2000-282523/24.
XX DR N-PSDB; A12975.
XX
XX Polynucleotide encoding human prostate-associated protease useful for
XX diagnosing and treating cancers; prostate disorders and
XX gastrointestinal disorders
XX
XX Claim 1; Fig 1A-B; 27pp; English.
XX
XX This sequence represents human prostate-associated protease (HUPAP). cDNA
XX encoding HUPAP was initially identified in a spinal cord cDNA library.
XX the cDNA encoding this sequence representing a consensus of overlapping
XX and/or extended nucleic acid sequences from spinal cord, prostate tumour
XX and colon cDNA libraries. HUPAP is a serine protease with structural and
XX functional homology with bovine enterokinase, human pancreatic kallikrein
XX and African rat renal kallikrein, sharing 38% homology with bovine
XX enterokinase. In addition, HUPAP is related to prostate-specific antigen
XX (PSA), a kallikrein which is a highly sensitive marker for prostate
XX cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
XX prevention and treatment of gastrointestinal disorders such as ulcerative
XX colitis, pancreaticitis, cancers, and prostatic disorders. HUPAP agonists
XX may be used to treat gastrointestinal disorders, and HUPAP antagonists
XX and inhibitors may be used to suppress excessive cell proliferation,
XX which is of use in cancer therapy. HUPAP nucleic acids are also useful
XX for generating hybridisation probes which may be used for mapping
XX naturally occurring genomic sequences.
XX
XX Sequence 283 AA:

```

```

Query Match 56.7%; Score 1540; DB 21; Length 283;
Best Local Similarity 99.6%; Pred. No. 4.2e-112;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 210 MKNTSAGNDIYKRLKHSACSKAVYSLRCIAGVNLNNSRSRIVGSGSALPGAMPW 269
DB 1 MKNTSAGNDIYKRLKHSACSKAVYSLRCIAGVNLNNSRSRIVGSGSALPGAMPW 60
OY 270 QVSLHVNQVHVGCGSITPEIVTAACVEKELNNPMHMTAFACILKSGFNYGAGTQVE 329
DB 61 qvslhvgnvhvcgsiltpeivtraahveklpnhpmhmtafagllqsfnyfagayqve 120
OY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLPNCAMLOPEOLCHTISWGATEE 389
DB 1 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLPNCAMLOPEOLCHTISWGATEE 389

```

```

DB 121 kvishpnysdsktkndialmkqlkpltfndlvkvcplpnhpmhmtafagllqsfnyfagatee 180
OY 390 KGKTSSEVLNAKAVLLIETORNSRYYDNLITPAMICAGFLGQNDVSCGDSGCLPYTSK 449
DB 181 kgktssevlnaakvllietqcnarsyydnllitpamlcagflgqnvdsqsgdxlvtak 240
OY 450 NNTWMLIGDTSKSGCAKATRPVYGVNMYFTDWITROMADG 492
DB 241 nntwmligdtswsgcakayrpvygvnmvftdwlyrqmzdg 283

```

RESULT 7

```

B32246
ID B32246 standard; Protein; 454 AA.
XX
XX B32246;
XX
XX 11-JAN-2001 (first entry)
XX
XX Tumour associated differentially-expressed gene 12 protein sequence.
XX
XX Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
XX tumour associated differentially-expressed gene 12; cytosolic; human;
XX malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
XX
XX Homo sapiens.
XX
XX MO200052044-A1.
XX
XX 08-SEP-2000.
XX
XX 02-MAR-2000; 2000MO-US05612.
XX PF
XX 03-MAR-1999; 99US-0261416.
XX PR
XX (UYAR-) UNTV ARKANSAS.
XX PA
XX O'Brien TJ, Underwood LJ;
XX PI
XX WPI; 2000-533263/48.
XX DR N-PSDB; A93842.
XX
XX DNA fragment encoding tumor associated differentially-expressed gene 12
XX protein used for diagnosing and treating malignant hyperplasia and
XX cancers including ovarian cancer
XX
XX Claim 3; Figure 4; 118pp; English.
XX
XX This invention relates to a novel transmembrane serine protease called
XX tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
XX located on chromosome 17. Sequences A93842-A93845 and B32246-B32249
XX represent human TADG-12 cDNA and their corresponding protein sequences. A
XX splice variant of TADG-12 (TADG-12v) leads to a truncated protein
XX product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12 exhibits
XX cytosolic activity, and can be used in vaccines and in gene therapy.
XX TADG-12 nucleotide and protein sequences are used in the diagnosis of
XX malignant hyperplasia and cancers of the ovary, breast, lung, colon,
XX prostate and other cancers where TADG-12 is overexpressed. TADG-12
XX proteins or fragments can be used to vaccinate an individual with cancer,
XX suspected of having a cancer or at risk of getting cancer. Sequences
XX A93846-A93853 represent PCR primers used for amplifying the TADG-12 cDNA
XX sequence, and in the quantitative analysis of TADG-12 mRNA. B32250
XX represents a peptide fragment of TADG-12, used to create anti-TADG-12
XX antibodies. Sequences B32251-B32369 represent TADG-12 peptides which
XX target HLA, and may be used in a vaccine or for immune stimulation.
XX
XX Sequence 454 AA:

```

```

Query Match 32.7%; Score 888; DB 21; Length 454;
Best Local Similarity 46.3%; Pred. No. 3.6e-61;
Matches 190; Conservative 52; Mismatches 148; Indels 20; Gaps 10;

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U09860; AAC50138.1; -
 CC EMBL: Y19124; CAB65555.1; -
 CC EMBL: Y19125; CAB65555.1; JOINED.
 DR EMBL: Y19126; CAB65555.1; JOINED.
 DR EMBL: Y19127; CAB65555.1; JOINED.
 DR EMBL: Y19128; CAB65555.1; JOINED.
 DR EMBL: Y19129; CAB65555.1; JOINED.
 DR EMBL: Y19130; CAB65555.1; JOINED.
 DR EMBL: Y19131; CAB65555.1; JOINED.
 DR EMBL: Y19132; CAB65555.1; JOINED.
 DR EMBL: Y19133; CAB65555.1; JOINED.
 DR EMBL: Y19134; CAB65555.1; JOINED.
 DR EMBL: Y19135; CAB65555.1; JOINED.
 DR EMBL: Y19136; CAB65555.1; JOINED.
 DR EMBL: Y19137; CAB65555.1; JOINED.
 DR EMBL: Y19138; CAB65555.1; JOINED.
 DR EMBL: Y19139; CAB65555.1; JOINED.
 DR EMBL: Y19140; CAB65555.1; JOINED.
 DR EMBL: Y19141; CAB65555.1; JOINED.
 DR EMBL: Y19142; CAB65555.1; JOINED.
 DR EMBL: Y19143; CAB65555.1; JOINED.
 DR EMBL: AL163218; CAB90392.1; -
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.156; -
 DR MIM: 226200; -
 DR InterPro: IPR000082; -
 DR InterPro: IPR000859; -
 DR InterPro: IPR000988; -
 DR InterPro: IPR001190; -
 DR InterPro: IPR001254; -
 DR InterPro: IPR002172; -
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF01390; SEA; 1.
 DR Pfam: PF00330; SRCR; 1.
 DR Pfam: PF000057; Ldl_recept_a; 2.
 DR Pfam: PF000089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00740; MAM_2; 1.
 DR PROSITE: PS01209; LdlRA_1; 2.
 DR PROSITE: PS50068; LdlRA_2; 2.
 DR Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
 KW Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT TRANSMEM 785 1019 CATALYTIC CHAIN (LIGHT CHAIN).
 FT DOMAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 182 223 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 225 334 CUB.
 FT DOMAIN 342 504 MAM.
 FT DOMAIN 524 634 CUB.
 FT DOMAIN 641 679 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 678 771 SRCR.
 FT ACT_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 876 876 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 971 971 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT LIPID 2 2 MYRISTATE (POTENTIAL).
 FT DISULFID 184 197 BY SIMILARITY.
 FT DISULFID 191 210 BY SIMILARITY.
 FT DISULFID 204 221 BY SIMILARITY.
 FT DISULFID 643 655 BY SIMILARITY.
 FT DISULFID 650 668 BY SIMILARITY.
 FT DISULFID 662 677 BY SIMILARITY.
 FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 810 826 BY SIMILARITY.

FT DISULFID 910 977 BY SIMILARITY.
 FT DISULFID 941 956 BY SIMILARITY.
 FT DISULFID 967 995 BY SIMILARITY.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 134 134 O -> E (IN REF. 3).
 FT CONFLICT 732 732 S -> P (IN REF. 3).
 FT CONFLICT 754 771 SOOGLDSTLRQCNHKS -> RRNAKKEIDALSPITLLA (IN REF. 3).
 SQ SEQUENCE 1019 AA; 112923 MW; B6AA245FD4A563 CRC64;

Query Match 23.8%; Score 647; DB 1; Length 1019;

Best Local Similarity 35.8%; Pred. No. 1.1e-40; Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

OY 113 CSNSGIECDSSGTCINPSMWCDSHCPGDEENRCVRLYG-----PNFLLQYSSQ 164
 DB 643 CKADHFQC-KNECVCPLVNLCDGLHCEGSDPADCVRFNGTNNGLVRRIO----- 696
 OY 165 RKSMBPVQDDMNENYGRAACRDMGKNNFYSSOGIVDGSSTFPRKLTLSAGNDIYRK 224
 DB 697 -SIMHTACAEENMTTQISNDVCCQLGLGSG-NSSKPIFSTDGC-PFVKLTMAP---DGLI 750
 OY 225 LYSNDCSSKAVYSLRC--IACGVNLSSR-QSRITGGSALPGAPMAYSLHYGVNHC 281
 DB 751 LTPSQCLDSTLRQCNHKSCKRKLAAODITPKYIVGSSNAEGAMPVWVGLYGGRLC 810
 OY 282 GGSIIIPENIVTAACVCRPLNPMHTAFAGILROSFNF--YGAGYQYKYSHPNYS 339
 DB 811 GASLVSSDMVSAHCVYGRNLEPSKWTAILGLHMSNLTSPQTVRLDELIVNHYNR 870
 OY 340 KTKNNDIALKLOKPLFENDLYKPVCLPNEGMLLOPEQLCWISGWCATEKGTSEVLA 399
 DB 871 RRKNDIAMHLEFKVNYDYIDIPCLPEENVFPFPRNCISLAGMCTVVYOGCTTANILDE 930
 OY 400 AKVLLLETRCNSRYVYDNLIPAMICAGFLQGVNDSOGDGGPVTYSKNINMILIGT 459
 DB 931 ADPPLLSNRCQOQMPREYN-TTENMICAGIEBGGIDSCGDSGGLPMLCENNRMFLAGVT 989
 OY 460-SWGSCKAKAYRPGVYGNVAVFTMDI 484
 DB 990 SFGYKCALPRRGVAVARSGFTFEMI 1014

RESULT 6
 ENTK_MOUSE
 ID ENTK_MOUSE STANDARD: PRT: 1069 AA.
 AC P97435:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```

FT DOMAIN 151 243 SRCR
FT ACT_SITE 236 492 CATALYTIC
FT ACT_SITE 236 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA; 53847 MM; 7EAF6DA18609DDA CRC64;

Query Match 99.2%; Score 2696; DB 1; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.3e-193;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNCGSPAIGYENHGYQENPYPAQFTVYEVHAPQYSPVQYAPRYLQA 60
DB 1 MALNCGSPAIGYENHGYQENPYPAQFTVYEVHAPQYSPVQYAPRYLQA 60
OY 61 SNPVCTQPKSPSGVCTSKTKALCITLTLGTFVLGALALMLKMGSKCSNGIEC 120
DB 61 SNPVCTQPKSPSGVCTSKTKALCITLTLGTFVLGALALMLKMGSKCSNGIEC 120
OY 121 DSGGTCTINSNMDGVSCHCGGEGDENRCVRLVGNPILQYSSORSKSMHVCDDMMYNT 180
DB 121 DSGGTCTINSNMDGVSCHCGGEGDENRCVRLVGNPILQYSSORSKSMHVCDDMMYNT 180
OY 181 GRACADMGKKNFYSSQGIYDSDGSTSEFKLNTSAGNDIYKRLYHSDACSSKAVSLR 240
DB 181 GRACADMGKKNFYSSQGIYDSDGSTSEFKLNTSAGNDIYKRLYHSDACSSKAVSLR 240
OY 241 CLACGVNLSSRSRIGVSGESALPGAMPQVSLHVONVHVGSGSIITPEVITPAHCVER 300
DB 241 CLACGVNLSSRSRIGVSGESALPGAMPQVSLHVONVHVGSGSIITPEVITPAHCVER 300
OY 301 PLNPNHMAFAFAGILQSPFHYGQYQVEVISHPNYDSTKKNNDALMLQPLTFNDL 360
DB 301 PLNPNHMAFAFAGILQSPFHYGQYQVEVISHPNYDSTKKNNDALMLQPLTFNDL 360
OY 361 VKRYVCLPNEGMLQPOLCISMGATEEKGKTESEVLANAKVLLIETQRCNSRYVDNLI 420
DB 361 VKRYVCLPNEGMLQPOLCISMGATEEKGKTESEVLANAKVLLIETQRCNSRYVDNLI 420
OY 421 TPAMICAGFLQGNVDSQCGDGGPPLVTSKNNIWMLIGDTSWGSQCAKAPRGYGVNMF 480
DB 421 TPAMICAGFLQGNVDSQCGDGGPPLVTSKNNIWMLIGDTSWGSQCAKAPRGYGVNMF 480
OY 481 TDWYIROMRADG 492
DB 481 TDWYIROMRADG 492
DB 481 TDWYIROMKANG 492

RESULT 2
TMS2_MOUSE STANDARD: PRT; 490 AA.
ID TMS2_MOUSE
AC 09J108; 09JUK4; 09QY82;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEIN, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
GN TMPSR52.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPSR52 in mouse and
human tissues."
RN J. Pathol. 191:0-0(2000).
RP [2]
RA SEQUENCE FROM N.A.
RC Han J., Kim S.;
RT "Putative transmembrane protease X."
RN Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RP [3]
RC SEQUENCE FROM N.A.
RA STRAIN-BALB/C;
RC Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin."
RN Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF199362; AAF97867.1; -
CC EMBL: AF243500; AAF64186.1; -
CC EMBL: AF113596; AAF21308.1; -
CC HSSP: P00761; IAKS.
CC MGD: MG1:1354381; TMPSR52.
CC InterPro: IPR001254; -
CC InterPro: IPR001314; -
CC InterPro: IPR002172; -
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01209; LDLRA_1; 1.
CC PROSITE: PS0068; LDLRA_2; 1.
CC Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
CC FM CYTOPLASMIC (POTENTIAL).
CC KW SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
CC TRANSMEM 84 104
CC FT 84 104
CC FT DOMAIN 105 490
CC FT DOMAIN 111 149
CC FT DOMAIN 150 242
CC FT DOMAIN 254 490
CC FT ACT_SITE 294 294
CC FT ACT_SITE 343 343
CC FT ACT_SITE 439 439
CC FT ACT_SITE 439 439
CC FT DISULFID 112 125
CC FT DISULFID 119 138
CC FT DISULFID 132 147
CC FT DISULFID 243 363
CC FT DISULFID 279 295
CC FT DISULFID 408 424
CC FT DISULFID 435 463
CC FT CARBOHYD 111 111
CC FT CARBOHYD 212 212
CC FT CARBOHYD 474 474
CC FT CONFLICT 122 122
CC FT CONFLICT 178 178
CC FT CONFLICT 320 320
CC FT CONFLICT 474 474
CC FT CONFLICT 474 474
CC FT SEQUENCE 490 AA; 53479 MM; 07D2B03EAD8A1A9 CRC64;

Query Match 77.9%; Score 2117; DB 1; Length 490;

```

```

Query Match      33.2%; Score 875.5; DB 1; Length 454;
Best Local Similarity 45.9%; Pred. No. 4,2e+58;
Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11;

QY LTLGTEFLVGAALAAIAQLKLFMGSKCSNSGIECDSSGTCINPNSMCDGVSHCGEDENRC 148
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db IIVIGIALIILALAIGLGIHF---DCSGK-YRGRSSFKEICELLARDGVSDCKDGEDEYKC 107
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY VRLYGPNFILOYVSSQRKSMHPVCODDMNENTGRACRDMGYKNFNYSQGIVDSGSTS 208
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db VRVGQNNAVLYQVFTA-ASMKTKMSDDMKGHYANVAACOLGFP-SYVSADNLRVSVLSLEGO 164
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY FMKLNTSAAGNV---DIYKKLHYS-----DACSSKAVYSRLCIACGVNINLSRDSRIYGBGS 261
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db FREEFVSIDHLIPDKKVATALHSYVYREGCAAGHVHTTLOCTAAGHRRGYS--SRIVYGNM 222
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY ALPGAMPQOVSLHVONNVHYCGSITTPETPIYVAACHVEKPLNNPMHWTFAGILROSEMF 321
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db SLSQMPMQASIQFGYHLCCGSVTPLMTITTAACV-YDLTLPMSWTIOVLV--SLLD 279
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY YGA-GYOEKYVISHPNYDSKTKNDINDIALMLQKPLTFENDLVKPVCLPNFGMKLQPEQLCW 380
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db NPAPSHLVEKIYHSHSKYPKRRLGNDIALMLAKLAGPLTFENMIQPVCLPNSEENPPDGKVCW 339
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY ISGCATEE-KGKTSEVUNAAKVALILEORCHRSRYVDNLITPAIACGFLGGNDSCOG 439
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db TSGCATATEDGAGDASPVLNHAAVPISIKICHNRDHYGGIIISPIMCACYLTGVDGSSCG 399
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY DSGGPLTVSKNNIMPLIGDTSMGSCAKARPGVGVNVAFTDMJYRQRRAD 491
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db DSGGPLVCOERLMKLVGATSFGICAEVKNKGVTTRYTSFLDMIHHEOMERD 451
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
ENTRK_BOVIN ID ENTRK_BOVIN STANDARD: PRT: 1035 AA.

AC PG8072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRESS OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
[1]
RN RN
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:7586-7592(1994).
[2]
RN RN
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehentulla A., Racie L.A., DiBlasio E.A.,
RA Ferenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RT J. Biol. Chem. 268:23311-23317(1993).
[3]
RN RN
RP SEQUENCE OF 801-827.
RC TISSUE-Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janaka H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RT J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRIPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
```

CC	-1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC	-1- TISSUE SPECIFICITY: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC	-1- PPM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC	-1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
DR	EMBL; U09859; AAB40026.1; .
DR	EMBL; L19663; AAA16035.1; .
DR	PIR; A61436; A61436.
DR	HSSP; P00763; IDPO.
DR	MEROPS; S01.156; .
DR	InterPro; IPR000082; .
DR	InterPro; IPR000859; .
DR	InterPro; IPR000998; .
DR	InterPro; IPR001190; .
DR	InterPro; IPR001254; .
DR	InterPro; IPR001314; .
DR	InterPro; IPR002172; .
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00629; MAM; 1.
DR	Pfam; PF01390; SEA; 1.
DR	Pfam; PF00530; SRCR; 1.
DR	Pfam; PF00057; Idl_recept_a; 2.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PROSITE; PS00134; TRYPsin_HIS; 1.
DR	PROSITE; PS00135; TRYPsin_SER; 1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS0060; MAM_2; 1.
DR	PROSITE; PS01209; LDLRA_1; 2.
DR	PROSITE; PS00068; LDLRA_2; 2.
KM	Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KV	Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
FT	CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT	CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
FT	TRANSSEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
FT	DOMAIN 240 350 CUB.
FT	DOMAIN 358 520 MAM.
FT	DOMAIN 540 650 CUB.
FT	DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
FT	DOMAIN 694 787 SRCR.
FT	ACT_SITE 841 841 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 892 892 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID 2 2 MYRISTATE (POTENTIAL).
FT	DISULFID 199 212 BY SIMILARITY.
FT	DISULFID 206 225 BY SIMILARITY.
FT	DISULFID 219 236 BY SIMILARITY.
FT	DISULFID 659 671 BY SIMILARITY.
FT	DISULFID 666 684 BY SIMILARITY.
FT	DISULFID 678 693 BY SIMILARITY.

OY 413 RYVDNLITPMICAGFLQGVNDSCGDSGGLPYTSKNNIMLIGDTSVSGCAKARPG 472
 Db 354 ddaygsevtckmccaglpbgvdtcgcgsdgpimysdq-whvvgivawgygcgspetcp 412
 OY 473 VYGNMVFETDIYROMRAD 491
 Db 413 VYKVSAYLIMVYVWKAe 431
 RESULT 13
 B08950
 ID B08950 standard; Protein; 480 AA.
 XX B08950;
 AC B08950;
 DT 30-AUG-2000 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.
 XX
 KW Human; secreted protein; cytostatic; anti-proliferative; vulnery;
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
 KW hyperproliferative disorder; infectious disease; tissue regeneration;
 KW screening; food additive; preservative; wound healing;
 KW hyper-vascular disease.
 XX
 OS Homo sapiens.
 XX
 PN MO200017222-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 22-SEP-1999; 99WO-US22012.
 XX
 PR 23-SEP-1998; 98US-0101546.
 PR 02-OCT-1998; 98US-0102895.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, N1 J;
 PI Komatsoulis G, Endress GA, Soppet DR;
 DR WPI: 2000-283538/24.
 XX
 PT Human secreted proteins and coding sequences useful in diagnostic and
 PT therapeutic methods for disorders such as immune system or
 PT proliferative disorders, related to the proteins
 PS Disclosure; Page 65; 416pp; English.
 XX
 CC The polynucleotide sequences given in A39052 to A39088 encode the human
 CC secreted proteins given in B08891 to B08984. The human secreted proteins
 CC can have activities based on the tissues and cells they are expressed in.
 CC Examples of the activities are: cytostatic; anti-proliferative;
 CC immunosuppressive; antibacterial; and vulnery. The secreted proteins
 CC and their related polynucleotide sequences are useful for diagnostic and
 CC therapeutic methods useful for diagnosing and treating disorders related
 CC to the secreted proteins. The proteins, and polynucleotide sequences may
 CC be useful for treating disorders of the immune system, hyperproliferative
 CC disorders, infectious disease, regeneration of tissues, for chemotaxis
 CC and for screening molecules that bind to the proteins. The proteins or
 CC polynucleotide sequences may be used as food additives or preservatives,
 CC to increase or decrease storage capabilities, fat content, lipid,
 CC protein, carbohydrate, vitamins, minerals, co-factors or other
 CC nutritional components, agonists or antagonists of the proteins may be
 CC used to prevent scar tissue growth during wound healing, and
 CC hyper-vascular diseases. A39043 to A39051 and B08890 are sequences used
 CC in the exemplification of the present invention.
 CC
 XX Sequence 480 AA;
 SO

Query Match 24.5%, Score 666.5; DB 21; Length 480;

Best Local Similarity 34.8%, Pred. No. 6,3e-44;
 Matches 157; Conservative 60; Mismatches 189; Indels 45; Gaps 12;
 OY 57 LTOASNPVCTOPKSPSGVCTSKTKALITLIGFLVAGLAAGLMMKMGSKSN 115
 Db 45 lsgacpc-aaslpvstfgavcmrigrcaavlglgyl-llegayvswllvlylcpaasqp 102
 OY 116 -SGIECDSSGTCINPSMKDCGVSHCPGGEDEKRCV-----RLGPNFILOYYSOR 165
 Db 103 lsgtlqdeeltl-----scseasaeaalpajpkvstfrinsedfllaqyrdq 151
 OY 166 KSMHPVODDNNENYGRACRDMGY-KNMFYSOGIVDDSGSTFMLNTSAGVNDLYKK 224
 Db 152 pwallvcnhevspalqlqclwslghlrlthkgyvlcd-----lknsgqefaqdspr 204
 OY 225 L-YHSDA-----CSSKAVVSLRCIACGVNLNSSROSRIVGESALPGAMPQVSLHVQ 276
 Db 205 lgfleeawgprinctagqvslrcsegar---plastrlvgsqavpgrvwpwgsvalg 261
 OY 277 NVHVGGSITPENVITVAHCEVK-PLNNPWHMTAFAGILNQSFMTGAGQVVKVISHP 335
 Db 262 fhncgsvvlprvvtahcmhbfirlarlsawrvhaglvhsavtrphgvalverllphp 321
 OY 336 NYDSKTKNDIAIMKIQKPLTFNDLYKPVCLPNPGMMLQPEQLCMISGWMGATEEKG-TS 394
 Db 322 lysqnhdydvalllrlqalnfstdvgaclpakqhfpxgrcwgvgntchpshtys 381
 OY 395 EVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGVNDSCGDSGGLPYTSKNNIMW 454
 Db 382 dmlqdtvplfstqlcnsscyygsaliprmicayldgradaagdsdgpvlvcpdgtwr 441
 OY 445 LIGDTSMGSCGCAKAVRPGVGNMVFETDIY 485
 Db 442 lvgvswgrgaephpgvyakvaeildwih 472
 RESULT 14
 B11698
 ID B11698 standard; Protein; 445 AA.
 XX
 AC B11698;
 XX
 DT 23-OCT-2000 (first entry)
 XX
 DE Mouse serine protease BSSP2 (mBSSP2), SEQ ID NO:8.
 XX
 KW BSSP2; serine protease; mouse; mBSSP2; human; mBSSP2; brain;
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;
 KW epilepsy; cancer; inflammation; inflammatory disorder; infertility;
 KW prostatic hypertrophy.
 XX
 OS Mus sp.
 XX
 PN MO200031272-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 19-NOV-1999; 99WO-JP06475.
 XX
 PR 20-NOV-1998; 98JP-0347785.
 PA (FUSO) FUSO PHARM IND LTD.
 PI Uemura H, Okui A, Komlaml K, Yamaguchi N, Mitsui S;
 DR WPI: 2000-400082/34.
 DR N-PSDB; A61662.
 XX
 PT Serine protease BSSP2, useful in detecting homologs, mutants and
 PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's
 PT disease, cancer, inflammation and prostate hypertrophy, using blood,
 PT urine or other tissues

A>Note: the sequence shown follows the authors' translation
R:Fujikawa, K.; Chung, D.W.; Hendrickson, L.E.; Davie, E.W.
Biochemistry 25, 2417-2424, 1986
A>Title: Amino acid sequence of human factor XI, a blood coagulation factor with four ts
A:Reference number: A00920; MID:86243360
A:Accession: A00920
A:Molecule type: mRNA
A:Residues: 1-625 <FUJ>
A:Cross-references: GB:M13142; NID:q182832; PID:AAA52487.1; PID:g182833
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2056-2060, 1991
A>Title: Location of the disulfide bonds in human coagulation factor XI: the presence of
A:Reference number: A37940; MID:91152017
A:Accession: A37940
A:Molecule type: protein
A:Residues: 38-333-49, 'X', 51-55, 'X', 57-63, 70-75, 'X', 77-79, 107-109, 'X', 111-112, 132-139,
'X', 280-282, 'X', 284, 285-297, 313-316, 'X', 318-319, 320-326, 'X', 328-330, 'X', 347-349, 373, 'X', 375
C:Comment: The proenzyme consists of two identical chains linked by one or more disulfide
he active site, and a heavy chain, which associates with high molecular weight (HMW) kin
C:Genetics:
A:Gene: GDB:Fl1
A:Cross-references: GDB:119891; OMIM:264900
A:Map position: 4q35-4q35
A:Insertions: 19/1; 73/2; 109/1; 162/2; 199/1; 252/2; 289/1; 343/2; 379/1; 435/2; 494/1; 52
C:Function:
A:Description: catalyzes the proteolytic activation of coagulation factor IX
A:Pathway: blood coagulation intrinsic pathway
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; glycoprotein; hemophilia C; homodimer; hydro
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-387/Product: coagulation factor XIa heavy chain #status experimental <HCH>
F:109-198/Domain: apple repeat <AP1>
F:109-198/Domain: apple repeat <AP2>
F:199-288/Domain: apple repeat <AP3>
F:390-379/Domain: apple repeat <AP4>
F:388-625/Product: coagulation factor XIa light chain #status experimental <LCH>
F:388-618/Domain: trypsin homology <TRY>
F:20-103, 514-581, 571-599/Disulfide bonds: #status predicted
F:29/Disulfide bonds: Interchain #status experimental
F:46-76, 50-56, 110-133, 136-165, 140-146, 200-283, 226-255, 230-236, 291-374, 317-346, 321-327, 38
F:80, 126, 353, 450/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:339/Disulfide bonds: Interchain #status predicted
F:387-388/Cleavage site: Arg-Ile (coagulation factor XIa) #status experimental
F:431, 480, 575/Active site: His, Asp, Ser #status predicted
F:491/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 19.6%; Score 533; DB 1; Length 625;
Best Local Similarity 30.5%; Pred. No. 4,4e-31;
Matches 122; Conservative 63; Mismatches 129; Indels 86; Gaps 11;

QY 169 HPC-----QDMNENYGRAACDMGKNNFVSOGIVDOSTSMKNTSGNDIY 222
DB 233 HPCLEFTFFSOEWPKEORNLCLKTSEGLPSTR--IKSKALSGPSLOCSHSHIPVF 290
QY 223 --KLYHS-----DACSSKAVYSLRCI-----ACGVNL 248
DB 291 CHSFFVHDTDFLEELDIYAAKSHACQKLTNAVRCQFTYTPAASCNGKCKVLT 350
QY 249 NSSR-----QSRVIGESALPAMPQVSLAVQNV 278
DB 351 SSNGSPKILHKGKISGYTLRLCKMDECTTKIKPRIVGTAASVREMPQVYLTHTSP 410
QY 279 ---HVCGSITTEWMTAHCVEKPLNPMHMTAFAGILROSPM-----FYGAGYQVEK 330
DB 411 TORHLCGGSTIGNQWILTAHCF-YGVESPIRLRVYSGLINQSEIKKDTSEFG---VQE 465
QY 331 VISHPNYSCTKNNDLAMKLRPLTFENDLVKPYCLPNCMMLOPEOLCWSIGWGATEEK 390
DB 466 IITHDYKMAESGVDIALLTLETIVNTDSORPICLPKSGDRNVIYIDCWYTGNGYKRLR 525
QY 391 GKTSVNLAAKVLLETRONRSRYVDNLTTPANICAGFLQGNVDSOGSGGGLVTSKN 450

DB 526 DKIONTLQAKKIPLYVNEEOCKR-RGKHTHKWICGIEGCKADKDGSGPLSCKHN 584
QY 451 NIMWLIDTSMGSGCAKAYRPGYGNMVTWYROMRA 490
DB 585 EVMHLVITSMGEGCAQRERPGYTNVVEVDILENTQA 624

RESULT 10
S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 06-Jan-1995 #sequence, revision 06-Jan-1995 #text, change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A>Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase
A:Reference number: S33777; MID:93305733
A:Accession: S33777
A:Molecule type: mRNA
A:Status: preliminary
A:Gene: preliminary
A:Residues: 1-416 <PAR>
A:Cross-references: EMBL:X70900; NID:q57928; PID:CAA50256.1; PID:q57929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TMN>
F:162-399/Domain: trypsin homology <TRY>
F:187-203, 290-358, 321-337, 348-380/Disulfide bonds: #status predicted
F:202, 256, 352/Active site: His, Asp, Ser #status predicted

Query Match 19.5%; Score 529.5; DB 1; Length 416;
Best Local Similarity 30.6%; Pred. No. 4,9e-31;
Matches 132; Conservative 66; Mismatches 177; Indels 57; Gaps 10;

QY 77 CTGKTKALCITITLGTFLYGALAAGLMKFMGSKNSGICEDSSGCTINPNCDDY 136
DB 11 CCSRPKVA---ALTGGLT-----FL-----TGGAASVAIVT 40
QY 137 SHCGGDENRCAVLYPNFLLAYSSQKSMHPCVODDMNENYGRAACDMGKNNFYS 196
DB 41 ILRSDPEPLXQVLSGDSRLVLDTBEETWRLCSSRANAVAGCGCEMFLRALH 100
QY 197 SQGIVDOSTSEFKLNTSA-----GNVDIYKLYHSDA---CSSKAVYSLPCICGVN 247
DB 101 SELDVRTAGAN-----GTSGFFCVDEGGLPLAORLLDIVISVCCPGRFETACODCG-- 153
QY 248 LNSRSRRIYGESALPGAMPQVSLHVQNVHCGSIIIPETIVTAARVERPLNPMH 307
DB 154 RRLPLVDRIYGGDDSLGRMPQVSLRYDGTHLGGSLSDGDLVLTARCFPERNRVLSR 213
QY 308 WTAFAGILROSPMEYGAQYQVEKVIYSHPNY-----DSKRNNDIALMKLRPLTFENDLY 361
DB 214 WRFPAGAVARTSP-HAVQVQAVIYHGILPRDRPIDNSNDIALVHSSSLPLEYI 272
QY 422 PAMTACAGFLQGNVDSOGSGGGLY-----TSKNNIMWLIDTSMGSGCAKAYRPGYGNV 477
DB 333 PKMFCAGYPPGGIDACGDSGGHGFVECDRIISGRNRLCGIVSMGTCALARRPGYTVR 392
QY 478 MVFTDRIYROMR 489
DB 393 IDREFWIFQAIK 404

RESULT 11
PLMS
Plasma (EC 3.4.21.7) precursor - mouse
N:Contains: angiotensin; plasminogen
C:Species: Mus musculus (house mouse)
C>Date: 20-Sep-1991 #sequence, revision 01-Nov-1996 #text, change 18-Jun-1999

Tue May 8 11:56:55 2001

us-09-323-597b-2.rag

Page 13

OY		402 VLLLEETORCSNRYVYNLITIPAMICAEFLGANDSCGGSGSGLPMTSKNIIMWLIGDTS 461 ::: :
Dδ	712 vpllnekccqcmpeyn-llcnmwcacaeysgavdsccgsdsgplmcgcennrvllaaystsf 770 ::	
OY		462 GSSCAKAVRPQGVGNNAVTFWDI 484 :
Dδ	771 gylcalpnurpgvyarvrficwml 793 :	

Search completed: May 8, 2001, 11:55:26
Job time: 46 sec

Oy	332	YGA-GYOVERKISHPNPDSDTKNNNDALMLQPLFENLIVRVCVLPJNNGMLOEQLCM	380
Db	280	npaephvkvkyvshykkykrrlgndalmlklsplifnmdipvcplnseenifpdkxw	339
Oy	381	ISGNGATEEKGKTSVLENAKVLIIETQRNSRYVDNLITPATICAGFLQGNVDSOCD	440
Db	340	tsygateedgdsaaipvlnhaavplsnklnchrdyvgllspmlcaagyltgvdascqgd	399
Oy	441	SGGFLVSKNNIMWLIGDITSMGSCAKARYPGYGVNVMFTMIYQGMAD	491
Db	400	sggplvceqerrlmkvlvatsfglgaevnkpvytrvtstfldwlnqmerd	450
RESULT	10		
ID	Y72093	Y72093 standard. Protein: 327 AA.	
AC	XX	Y72093;	
XX	XX	28-MAR-2001 (first entry)	
DE	XX	Human serine protease #4 encoded by clone HCHAK72.	
KW	XX	Human; serine protease; osteopathic; immunosuppressive; anti-allergic;	
KW	XX	anti-inflammatory; cytostatic; cardiant; neuroprotective; nootropic;	
KW	XX	neuroleptic; vulnary; ophthalmological; antibacterial; antiviral;	
KW	XX	antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma	
KW	XX	treatment; bone formation disorder; osteoporosis; arthritis; cancer;	
KW	XX	connective tissue disorder; autoimmune disorder; wound healing; asthma;	
KW	XX	systemic lupus erythematosus; male reproductive system disorder;	
KW	XX	testicular cancer; digestion and food absorption disorder; arrhythmia;	
KW	XX	Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;	
KW	XX	behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;	
KW	XX	cardiovascular disorder; ocular disorder; drug screening.	
XX	OS	Homo sapiens.	
XX	XX		
PN	XX	WO200068247-A2.	
PD	XX		
PD	XX	16-NOV-2000.	
PF	XX	05-MAY-2000; 2000MO-US12207.	
PR	XX	07-MAY-1999; 99US-0133239.	
PR	XX	20-MAY-1999; 99US-0135163.	
PR	XX	03-AUG-1999; 99US-0147005.	
PR	XX	09-SEP-1999; 99US-0152935.	
PR	XX	01-NOV-1999; 99US-0162979.	
PA	XX	(HUMA-) HUMAN GENOME SCI INC.	
PI	XX		
PI	XX	Ruben SM, Shi Y, Young PE, Ni J.	
XX	XX		
DR	XX	WPI: 2000-679799/66.	
DR	XX	N-PSDB: D02323.	
PT	XX	New nucleic acid molecules encoding human serine protease polypeptides,	
PT	XX	useful for diagnosis, prevention and/or treatment of disorders e.g.	
PT	XX	osteoporosis, lupus erythematosus and Alzheimer's -	
PS	XX	Claim 12; Page 280-281; 289pp; English.	
XX	XX		
CC	CC	The present sequence is human serine protease #4 from clone	
CC	CC	HCHAK72 (ATCC Deposit No. PRA27).	
CC	CC	The invention relates to human serine proteases and their cDNA clones.	
CC	CC	It is used in methods for the diagnosis, prevention and treatment of	
CC	CC	various disorders related to serine protease such as bone formation	
CC	CC	disorders (osteoporosis), connective tissue disorders (arthritis),	
CC	CC	autoimmune disorders (systemic lupus erythematosus), wound healing, male	
CC	CC	reproductive system disorders (testicular cancer), digestion and food	
CC	CC	absorption disorders (Crohn's disease), neurodegenerative diseases	
CC	CC	(Alzheimer's disease), behavioral disorders (Tourette's syndrome),	
CC	CC	(Alzheimer's disease), behavioral disorders (Tourette's syndrome),	

CC	proliferative and cancerous conditions (acute myelogenous leukemia),
CC	allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC	ocular disorders (glaucoma) and infectious diseases caused by bacteria,
CC	viruses, fungi or parasites. It is also useful for screening therapeutic
CC	compounds. Serine proteases are used as immunological probes or
CC	polymorphic markers for the identification of chromosomes, cells and
CC	tissues in biological samples, identification of male contraceptive
CC	agents, delivery of compositions to targeted cells expressing a
CC	receptor for serine protease, hybridisation probes and molecular weight
CC	markers. Serine protease nucleic acids are also useful in gene therapy.
CC	Note: The present sequence shown in page 280-281 of sequence listing has
CC	been assigned SEQ ID NO: 14. But the sequence, human serine protease
CC	epidermal growth factor (EGF)-like domain (Y7211a) shown in
CC	page 12 is also referred as SEQ ID NO: 14.
CC	
XX	Sequence 327 AA:
XX	
Query Match	27.3%; Score 742.5; DB 21; Length 327;
Best Local Similarity	46.7%; Pred. No. 4.9e-50;
Matches 134; Conservative	45; Mismatches 116; Indels 15; Gaps 8;
QY	171 VCDNNENYGRACRADMGYKNKFSSOGIVDDSGSTFSMKLNTSAGNV--DIYKLYH 227
Db	1 mcsdckkyhyanvaceagqifp-syvsdnllvyslsqgffreefsidhllpddkvtcalnh 59
QY	228 S---DAGSKAVYSLRCIACGVNLNNSKSHVGSSEALPGANPMQVSLHYNVHYCG 283
Db	60 syvrgcasghvnlclqctacghrzys--grlygmmalslsqpwqaslsqfgyhllcgg 117
QY	284 SITPEWITYAHCEKPLNNPWHMTARGLIRGSFMYGA-GIQVEKVIASHPNYDSKTR 342
Db	118 svtlpwlitlaahcv-ydilylpkswtlcvgylv--slldnpapshlveklyhskypkxl 174
QY	343 NNIDIAMLKQKPLTFENDLVKPYCLNPGMVLQPEQLCWISGMGATEE-KGKTSEVLAIAK 401
Db	175 gndlaamlkagpltfliemqgvclpnsseenfpdqkvctswgctegadgaapvlnhaa 234
QY	402 VLLIETORCNSRYVDNLITTPAMICAGFLOGNVDSQCGDSGGLPVTSKNNIMWLGDTSM 461
Db	235 vplnskhkcnhdvysglispsmlaaglyltsgvdsccgdsqgprlvqgerllwklvgatgf 294
QY	462 GSGCAKAYRPBGVYGNVWFTDWTYQMRAD 491
Db	295 glgcaevnkpvytrvtsfldwlhagmeard 324

	PR	16-SEP-1998;	98US-0100664.
	PR	17-SEP-1998;	98US-0100683.
	PR	17-SEP-1998;	98US-0100684.
	PR	17-SEP-1998;	98US-0100710.
	PR	17-SEP-1998;	98US-0100711.
	PR	17-SEP-1998;	98US-0100919.
	PR	17-SEP-1998;	98US-0100930.
	PR	18-SEP-1998;	98US-0100848.
	PR	18-SEP-1998;	98US-0100849.
	PR	18-SEP-1998;	98US-0101014.
	PR	18-SEP-1998;	98US-0101068.
	PR	18-SEP-1998;	98US-0101071.
	PR	22-SEP-1998;	98US-0101279.
	PR	23-SEP-1998;	98US-0101471.
	PR	23-SEP-1998;	98US-0101472.
	PR	23-SEP-1998;	98US-0101474.
	PR	23-SEP-1998;	98US-0101475.
	PR	23-SEP-1998;	98US-0101476.
	PR	23-SEP-1998;	98US-0101477.
	PR	23-SEP-1998;	98US-0101479.
	PR	24-SEP-1998;	98US-0101738.
	PR	24-SEP-1998;	98US-0101741.
	PR	24-SEP-1998;	98US-0101743.
	PR	24-SEP-1998;	98US-0101915.
	PR	24-SEP-1998;	98US-0101916.
	PR	29-SEP-1998;	98US-0102207.
	PR	29-SEP-1998;	98US-0102240.
	PR	29-SEP-1998;	98US-0102307.
	PR	29-SEP-1998;	98US-0102330.
	PR	29-SEP-1998;	98US-0102331.
	PR	30-SEP-1998;	98US-0102484.
	PR	30-SEP-1998;	98US-0102487.
	PR	30-SEP-1998;	98US-0102570.
	PR	30-SEP-1998;	98US-0102571.
	PR	01-OCT-1998;	98US-0102684.
	PR	01-OCT-1998;	98US-0102687.
	PR	02-OCT-1998;	98US-0102965.
	PR	06-OCT-1998;	98US-0103258.
	PR	06-OCT-1998;	98US-0103449.
	PR	07-OCT-1998;	98US-0103314.
	PR	07-OCT-1998;	98US-0103315.
	PR	07-OCT-1998;	98US-0103328.
	PR	07-OCT-1998;	98US-0103395.
	PR	07-OCT-1998;	98US-0103396.
	PR	07-OCT-1998;	98US-0103401.
	PR	08-OCT-1998;	98US-0103633.
	PR	08-OCT-1998;	98US-0103678.
	PR	08-OCT-1998;	98US-0103679.
	PR	08-OCT-1998;	98US-0103711.
	PR	14-OCT-1998;	98US-0104257.
	PR	20-OCT-1998;	98US-0104987.
	PR	20-OCT-1998;	98US-0105000.
	PR	20-OCT-1998;	98US-0105002.
	PR	21-OCT-1998;	98US-0105104.
	PR	22-OCT-1998;	98US-0105169.
	PR	22-OCT-1998;	98US-0105266.
	PR	26-OCT-1998;	98US-0105693.
	PR	26-OCT-1998;	98US-0105694.
	PR	27-OCT-1998;	98US-0105807.
	PR	27-OCT-1998;	98US-0105881.
	PR	27-OCT-1998;	98US-0105882.
	PR	28-OCT-1998;	98US-0106023.
	PR	28-OCT-1998;	98US-0106029.
	PR	28-OCT-1998;	98US-0106030.
	PR	28-OCT-1998;	98US-0106032.
	PR	28-OCT-1998;	98US-0106033.
	PR	29-OCT-1998;	98US-0106178.
	PR	29-OCT-1998;	98US-0106248.
	PR	29-OCT-1998;	98US-0106384.
	PR	30-OCT-1998;	98US-0108500.
	PR	30-OCT-1998;	98US-0108466.
	PR	03-NOV-1998;	98US-0106856.
	PR	03-NOV-1998;	98US-0106905.
	PR	03-NOV-1998;	98US-0106919.
	PR	03-NOV-1998;	98US-0106932.
	PR	10-NOV-1998;	98US-0106934.
	PR	10-NOV-1998;	98US-0107783.
	PR	17-NOV-1998;	98US-0108775.
	PR	17-NOV-1998;	98US-0108779.
	PR	17-NOV-1998;	98US-0108787.
	PR	17-NOV-1998;	98US-0108788.
	PR	17-NOV-1998;	98US-0108801.
	PR	17-NOV-1998;	98US-0108802.
	PR	17-NOV-1998;	98US-0108806.
	PR	17-NOV-1998;	98US-0108807.
	PR	17-NOV-1998;	98US-0108867.
	PR	17-NOV-1998;	98US-0108925.
	PR	18-NOV-1998;	98US-0108848.
	PR	18-NOV-1998;	98US-0108849.
	PR	18-NOV-1998;	98US-0108850.
	PR	18-NOV-1998;	98US-0108851.
	PR	18-NOV-1998;	98US-0108852.
	PR	18-NOV-19	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:54:40 ; Search time 20.73 Seconds

(without alignments)
1631.051 Million cell updates/sec

Title: us-09-323-597b-2

Perfect score: 2717

Sequence: 1 MALNGSPPAIGPYENHGY.....VYGVNMFPTDITRQMRADG 492

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries:

Database :

1: PIR67:*

2: PIR1:*

3: PIR3:*

4: PIR4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 JE0315	low-density lipopr
5	569.5	21.0	638	1 KQHUP	plasma kallikrein
6	565.5	20.8	417	1 S00845	hepsin (EC 3.4.21.
7	556	20.5	638	1 KORPPL	plasma kallikrein
8	538	19.8	638	1 KOMSPL	plasma kallikrein
9	533	19.6	625	1 KFHU1	coagulation factor
10	529.5	19.5	416	1 S33777	hepsin (EC 3.4.21.
11	518.5	19.1	821	1 PLM6	plasma (EC 3.4.21
12	501	18.4	421	1 S11674	acrosin (EC 3.4.21
13	501	18.4	1524	2 T30337	polyprotein - Afri
14	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
15	489.5	18.0	436	2 JX0172	acrosin (EC 3.4.21
16	486.5	17.9	761	2 JCS759	brain-specific ser
17	483	17.8	418	2 A37344	acrosin (EC 3.4.21
18	477	17.6	790	1 PLPG	plasma (EC 3.4.21
19	475.5	17.5	431	2 S47538	acrosin (EC 3.4.21
20	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
21	472.5	17.4	810	2 B30848	plasma (EC 3.4.21
22	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
23	470.5	17.3	810	2 I46260	plasma (EC 3.4.21
24	461	17.0	343	1 A57014	proctasin (EC 3.4.
25	456.5	16.8	460	2 B61545	plasma (EC 3.4.21
26	455	16.7	812	1 PLBO	plasma (EC 3.4.21
27	451	16.6	2616	2 A57096	nudel protein prec
28	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21
29	450	16.6	270	2 S56160	mast cell trypase

30	449	16.5	275	2 S40005	trypsin (EC 3.4.21
31	449	16.5	655	1 A46688	hepatocyte growth
32	448	16.5	277	2 S35340	trypsin (EC 3.4.21
33	447	16.5	276	2 A38654	mast cell proteina
34	446.5	16.4	420	2 A55283	acrosin (EC 3.4.21
35	445.5	16.4	810	1 PLHU	plasma (EC 3.4.21
36	444.5	16.4	455	2 A61545	plasma (EC 3.4.21
37	443.5	16.3	266	2 S54146	trypsin (EC 3.4.21
38	442.5	16.3	458	1 S00657	apoptotrin(a) (EC
39	441.5	16.2	786	1 A47547	trypsin (EC 3.4.2
40	441	16.2	274	2 A47246	serine proteinase
41	439	16.2	273	2 A47246	trypsin (EC 3.4.2
42	438	16.1	274	2 S35339	trypsin (EC 3.4.21
43	437	16.1	275	2 S40007	trypsin (EC 3.4.21
44	436	16.0	242	2 S49489	trypsin (EC 3.4.21
45	436	16.0	456	1 KXBO	protein C (activat

ALIGNMENTS

RESULT 1

A43090

enteropeptidase (EC 3.4.21.9) precursor - bovine

N:Alternate names: enterokinase

C:Species: Bos primigenius taurus (cattle)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A43090; A48874; A61436

R:Kilamoto, Y.; Yuan, X.; Mu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A:Title: Enterokinase, the Initiator of Intestinal digestion, is a mosaic protease co

A:Reference number: A43090; MUID:94329561

A:Accession: A43090

A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1035 <KIR>

A:Cross-references: GB:U09859; NID:9746410; PIDN:AAA0026.1; PID:9746411

A:Experimental source: small intestine

R:Lavallie, E.R.; Rehentulla, A.; Recler, L.A.; DiBlasio, E.A.; Ferez, C.; Grant, K.L

J. Biol. Chem. 268, 23311-23317, 1993

A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit o

A:Reference number: A48874; MUID:94043122

A:Accession: A48874

A:Molecule type: mRNA

A:Residues: 801-1035 <LAV>

A:Cross-references: GB:U19663; NID:9416131; PIDN:AAA16035.1; PID:9416132

A:Note: parts of this sequence, including the amino end of the mature protein, were c

R:Alight, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A:Reference number: A61436; MUID:92189715

A:Accession: A61436

A:Molecule type: protein

A:Residues: 801-807, 'Y', 809-827 <LIC>

C:Comment: The mechanism of association with the membrane of the intestinal brush bor

embrane attachment using a signal-anchor sequence.

C:Comment: Conversion from membrane-bound to soluble forms may involve further proces

C:Complex: mature enteropeptidase is variously reported to contain two (heavy and lig

liffe linked

C:Function:

A:Description: cleaves propeptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane prot

F:322-18/Domain: transmembrane #status predicted <TM>

F:32-117/Product: enteropeptidase mini chain #status predicted <KM>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HC>

F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:358-520/Domain: MAM homology <MAM>

F:442-647/Domain: Clr/Cls repeat homology <CLR>

F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:694-799/Domain: scavenger receptor cysteine-rich domain homology <status atypical <

F:801-1035/Product: enteropeptidase light chain #status predicted <LC>

F:801-1030/Domain: trypsin homology <TRY>
 F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
 F:788-912,826-842,926-933,957-972,983-1011/Disulfide bonds: #status predicted
 F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 660; DB 1; Length 1035;
 Best Local Similarity 33.9%; Pred. No. 5e-40;
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

65 VCTQPKSPSGTCTSKTKKALCITLT-----GPELVGAALAGLIMKFMKSCNSG 117
 611 VYRGPGVNVNFSTNMWTLFTIDNNLAKOGFRANFTTQGLG-----IPEPCEDN 663
 118 IECDSSSTCINPNCDCVSHCPGEGEDNRCVRLG-----PNTLYQVSSQKSMH 169
 664 FQC-KDECIPLVNLCDGPHCKDGSDEAHCVRLFNCTTDSGLVQFRIO-----SIWH 716
 170 PVODDMENYGRNACRDMGYKNNFYSSQGIYDSDGSTSPMKLNTSAGNDIYKLYHSD 229
 717 VACAEHTTQISDDVCOLLGLTG--NSVPTSTGGGPPVNLTAAGSLT---LTPSQ 771
 230 ACSSKAVVSLRC--IACGVNLNNSROS-RIVGESALPGAMPQVSLHVNHYCGSII 286
 772 QCLIEDSLILQCHYKSCGKLTQEVSPKIVGSDSREGAMPVVALYFDQVCGASIV 831
 287 TPEWYTAACVCEKPLNPNHMTAFAGILROSEMFYAGYQVE-----KYISHPNYDSKT 341
 832 SRMLVSAACHVGRNNEPSPKRAVLGLHNASNL--TSPQIEIRLLIDQYVNHYKRR 888
 342 KNNIDIALMKLOKPLTENDLVKPYCLPMPGMLOPEOLCMTISGMCATEBEKKTSEVLNAK 401
 889 KNNIDIAMHLEKVNITYDIOTICLPENOVFPFGKICSIAGKCALIYQGSTADYLDQAD 948
 402 VLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSCGSDSGGLPYTSKNNIMWILGDTSW 461
 949 VPLLSNERCQOQMPHYN-ITENMYCAGYEAGVDSGCGSGPLMCCENRMFLAGVTSE 1007
 462 GSGCAKAYRPGVYGNVAFETDWT 484
 1008 GYQCALPNRPGVYARVRFETEMI 1030

RESULT 2

enteropeptidase (EC 3.4.21.9) precursor - human
 N:Alternate names: enterokinase
 C:Species: Homo sapiens (man)
 C>Date: 19-MAY-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A56318; BA3090
 R:Altmeto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
 Biochemistry 34, 4562-4568, 1995
 A:Title: CDNA sequence and chromosomal localization of human enterokinase, the proteolyt
 A:Reference number: A56318; MUID:95234679
 A:Accession: A56318
 A:Molecule type: mRNA
 A:Residues: 1-1019 <KIT>
 A:Cross-references: GB:U09860; NID:9746412; PIDN:AAC50138.1; PID:9746413
 R:Altmeto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A:Reference number: A43090; MUID:94329561
 A:Accession: BA3090
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 749-1019 <K12>
 A:Cross-references: GB:U09860
 C:Comment: The mechanism of association with the membrane of the intestinal brush border
 (noted below) or with amino-terminal myristoylation of the heavy chain.
 C:Genetics:
 A:Gene: GDB:PRS57
 A:Cross-references: GDB:384083; OMIM:226200
 A:Map position: 21q21-21q21

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
 ducts.

C:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymoge

F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
 F:184-921/Domain: transmembrane #status predicted <TM>

F:184-921/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:526-631/Domain: NAM homology <NAM>
 F:526-631/Domain: Clr/Cls repeat homology <CLR>

F:678-783/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
 F:785-1019/Domain: trypsin homology <TRY>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin

F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
 F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 23.8%; Score 647; DB 1; Length 1019;

Best Local Similarity 35.8%; Pred. No. 4.3e-39;
 Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

113 CSNSIECDSSSTCINPNCDCVSHCPGEGEDNRCVRLG-----PNTLYQVSSQ 164
 643 CKADHQC-KNGECVPLVNLCDGHLCEGSDPADCVRFNGTNNGLVRRIO----- 696
 165 RKMHPPCODNMENYGRNACRDMGYKNNFYSSQGIYDSDGSTSPMKLNTSAGNDIYK 224
 697 -STWHRAEAEHTTQISDDVCOLLGLTG--NSKPLFTSDG-PEFKLNTAP---DGLI 750
 225 LYHSDACSSKAVVSLRC--IACGVNLNNSR-OSRIVGESALPGAMPQVSLHVNHYC 281
 751 LTPSQOCLDLSILRLQCHYKSCGKLTQADITPKIYGSNAGKAMPVVALYGGGRLLC 810
 282 GGSITPEWYTAACVCEKPLNPNHMTAFAGILROSEMFYAGYQVE-----KYISHPNYDSKT 339
 811 GASVSSDWLVSAACHVGRNNEPSPKRAVLGLHNASNL--TSPQIEIRLLIDQYVNHYKRR 870
 340 KTKNNIDIALMKLOKPLTENDLVKPYCLPMPGMLOPEOLCMTISGMCATEBEKKTSEVLNA 399
 871 RKNNDIAMHLEKVNITYDIOTICLPENOVFPFGKICSIAGKCALIYQGSTADYLDQAD 930
 400 AKVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSCGSDSGGLPYTSKNNIMWILGDT 459
 931 ADVPLLSNERCQOQMPHYN-ITENMYCAGYEBEGIDSCGSDSGPLMCCENRMFLAGVT 989
 460 SWGSCAKAYRPGVYGNVAFETDWT 484
 990 SEGYKCALPNRPGVYARVRFETEMI 1014

RESULT 3

enteropeptidase (EC 3.4.21.9) precursor - pig
 N:Alternate names: enterokinase
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A53663
 R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakai, N.; Tsukada, S.; Miki, K.; Kurokawa
 J. Biol. Chem. 269, 19976-19982, 1994
 A:Title: Structural characterization of porcine enteropeptidase.
 A:Reference number: A53663; MUID:94327548
 A:Accession: A53663
 A:Molecule type: mRNA
 A:Residues: 1-1034 <MAT>
 A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
 A:Note: parts of this sequence, including the amino ends of three chains isolated fro
 C:Comment: The mechanism of association with the membrane of the intestinal brush bor
 (noted below) or with amino-terminal myristoylation of the heavy chain.

C:Function: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolyase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat
C:Keywords: glycoprotein; hydrolase; serine proteolase; transmembrane protein; zymogen
F:12-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase min chain #status predicted <MCb>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCb>
F:199-336/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>
F:541-646/Domain: C1r/C1s repeat homology <C1r>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SR>
F:800-1034/Product: enteropeptidase light chain #status predicted <LCb>
F:800-1029/Domain: trypsin homology <TRY>
F:116-147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,988
F:787-911,925-841,925-932,956-971,982-1010/Disulfide bonds: #status predicted
F:860,891,986/Active site: His, Asp, Ser #status predicted

Query Match	23.3%	Score 634;	DB 1;	Length 1034;
Best Local Similarity	33.9e-38;	Pred. No. 3,9e-38;		
Matches 147;	Conservative 72;	Mismatches 179;	Indels 44;	Gaps 13;

QY	65	VCIOPKSPSGVCTSKTKKALCIT---	LTIG---	TLVGAALAAGLIMKFMGSKCSNSG	117
DB	610	VYTPPGPEVEDFSTTNRMTVLFITNDALTKGGFRANFTTGYHLG-----	IPECKEDN		662
QY	118	IECDSSGTCINPMSKCVSHCPGEGEDENRCRYLYG--PNFIIQVYSQSRKSMHPVCODD			175
DB	663	FGCE-NKECVLLVNLCDGFSHCCKDGSDEAHQVRLNGLNANNSGLVQFPIQSIIMHTACLEN			721
QY	176	WNEYGRACEDMGYKNN-----FYSSQGIYDSDGSGTSFMKLNTSAGAVDIYKKLYHSDA			230
DB	722	WTOTSDVVCOLLIGTGNSSMPEFS-----GGGGPFVPLNTAPNQSLLI---	LTASEQ		771
QY	231	CSKRAYVSLRC--IACQVNLNNSROS--RIYGGESALPCAMPQVSLYHOVNYHVGGSITT			287
DB	772	CFEDSLILLCNMHSCCKKOYAOEVSPIRIVGNDNSREDAWPMVYALVYINGOLLGASLVS			831
QY	288	PEWIVTAHCEYERPLANNPMHTAFAGILROSFMFYAGYQ-----VEKVIISHPNYSKTK			342
DB	832	RDMLVSAHACHYCGNLPSPSKKAILLGLHMTSNL--TSPQIVTRLDEIVINPHYNRARK			888
QY	343	NNDIATLAKLKLPLEFNDLVKPVCLPNPQMGLQPEOLCISGSGATEEKGKSEVYLANAKV			402
DB	889	DSDIAMMHLEKAVYTDIYQICLPEENQVFPFPRGICSIAGMKGYIYDGSADILLQEDAV			948
QY	403	LLIETORCNSRYVDNLITPAMICAGFLQGNVSDCOGDSGGPLVTSKNNIMWLIGDTSWG			462
DB	949	PLTSENKCOQOMPEYN-ITENMMCAGYEEDGIDSCOGDSGFLMCLLENRWLLAGVTSFG			1007
QY	463	SGCAKATRPQYGVNMYFTDNI			484
DB	1008	YQCALPNRPQVYATAPKFTEMI			1029

```

RESULT 4
JE0315
low-density lipoprotein receptor-related protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
C/Accession: JE0315
R/Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A/Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A/Reference number: JE0315, MUID:9842596
A/Accession: JE0315
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1113 <TOM>

```

A: Superfamilies: trypsin homology; LDL receptor ligand-binding repeat homology
C: Cross-references: DDBJ:AB013874; NID:q368914; PIRN:AAA4337.1; PID:q3689145
E: 337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
E: 374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
E: 410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
E: 447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
E: 648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
E: 684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
E: 723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
E: 869-1097/Domain: trypsin homology <TRY>

```

Query March 23.3%; Score 577.5; DB 2; Length 1113;
  .Best Local Similarity
Matches 134; Conservative 67; Mismatches 155; Indels 47; Gaps 14;

QY 111 SKCSNGSIECDSSGTCIIPMSWMCDSVSHCPGGEDENCVRLY--GPNFIQVYSSQKS 167
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 721 SFCDNNELEC-ANHECVPRDLCMDGVDCSSDSEMGCVTLTSKNSSSLLTVHKSAXE- 778

QY 168 WHPYCDQDMNNTYGAACRDMGYKNMFYSOGIYVDSGSTSPMKLTASGVND--ITYKK 224
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 779 -HAYCADQWRETLISLACKOMGLER--SVTKLIGQSGQOHLRIYPMENLNGSTLDEL 835

QY 225 LYHSDACSKAAYSLRCIA--CGVNLSSRSRIVGEGSALPGANPQVSLHWQ-IVHVC 281
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 836 LVYHRSCTSRSEISILSCSKDQGRPAARMKRIILGRTSRGPRWPMQCSLQSEPSGHIC 895

QY 282 GGSIIITPEWITYAAHCV-----KPLANNMHHTAFGILIRQSPMFQGAQY 327
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 896 GCVLIAKKVVLTVAHCFEGRBDADVMKYVFGIINNLDHP-----SGFMQTRP----- 941

QY 328 VEKYSIHRYNDSTKRNNDIALMLQKPLTFPNDLVKPYCLPFPNGMMLQEPOLCWMISGMAT 387
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 942 VKTILLHRYSRAYVDYDISVELSDOINERSYAPVCLSPREEVLEPDTYCIYIGWG-- 999

QY 388 EEKGTSEVYLAAKVLLIETORCNSRYVDNLITPAMICAGELQGVNVDSCGDSGGLPYT 447
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1000 HMGKNMPEKLOEGEVRILLPEOCOS-YEDMKITINRMICAGIESTGVDSQMSGGLPYVC 1058

QY 448 SK-NNIMWILIGTWSGSGC-AKAYRPGVYGVNVMTDITRYROM 488
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1059 ERPGGQMTLFGTLTSHGVSCTSKVLYGPGVYSNVSYVGIENKQI 1101

```

RESULT 5

plasma kallikrein (EC 3.4.21.34) precursor - human

N:Alternate names: Kininogenin; plasma prekallikrein

C:Species: Homo sapiens (man)

C>Date: 13-Aug-1986 #sequence_rev1ston 13-Aug-1986 #text_change 18-Jun-1999

C:Accession: A00921, A37939

C:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

A>Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four

A:Reference number: A00921; MUID:86243359

A:Accession: A00921

A:Molecule type: mRNA

A:Residues: 1-638 <CR>M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263

A:Cross-references: GB:U03333

R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A>Title: Location of the disulfide bonds in human plasma prekallikrein: the presence

A:Reference number: A37939; MUID:91152016

A:Accession: A37939

A:Molecule type: protein

A:Residues: 20-27,140-46, 'X', 48, 'H',50, 'X',52-70, 'H',75-76, 'X',78-80;103-113,131-140,1

;260-283, 'X',285;287-291, 'X',293-295;314-317, 'X',319-320;321-324, 'X',329-333;334-339,

525;538-551;552, 'X',564-567;573, 'X',575-576;578-583, 'X',585;592-604 <MCM>

C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex

C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a

are linked by one or more disulfide bonds.

C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciproca

hogen and may also play a role in the renin-angiotensin system by converting prorenin I
 C:Genetics:
 A:Gene: GDB:KLR3
 A:Cross-references: GDB:127575; OMIM:229000
 A:Map position: 4q35-4q35
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <Sto>
 F:20-638/Product: plasma kallikrein #status predicted <MAT>
 F:20-639/Domain: plasma kallikrein heavy chain #status predicted <HCH>
 F:20-109/Domain: apple repeat <AP1>
 F:110-139/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-639/Domain: plasma kallikrein light chain #status predicted <LCH>
 F:21-104/Domain: trypsin homology <TRY>
 F:21-104-47-77-51-57-111-194-137-166-141-147-201-284-227-256-231-237-292-375-322-328-383
 F:127-308-336-453-494/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:318-347-340-345/Disulfide bonds: #status predicted
 F:390-391/Cleavage site: Arg-1le (coagulation factor XIa) #status predicted
 F:434-483-578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 569.5; DB 1; Length 638;
 Best Local Similarity 32.8%; Pred. No. 1e-33;
 Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

55 RVLTQASNVNVTQTPSPGTVCTSKTKALCTLTGLFLVGAALAGLTKRMMS--- 111
 218 RVLT--PDAPVCC-----TICTYHPN--CLEFTFY-----NVMKIESQRNV 255
 112 ---KCSNSGIECDG-----SG---TCINP-----SNKCDVSHCPGGEENRCVRL 151
 256 CLKTKSESTPSSSTPQENTISGYSLTLCKRLPEPCSKITYGVD--GGELN----- 308
 152 YGPNFLQYSSORSKSWHPYCODDMNENTGRAACRDMGKNNFYSSGQYIDSGS--- 208
 309 --VTFVKGV-----NVCOE-----TCTRM-TRCOFFYSLPDECKEKCKC 347
 209 FPKLNMSAANDVIYKKLYHSDACSSKAVYSLRCIAGVN--LNSSQSRHYGGSALPGA 266
 348 FRLKMSDGSPTRI--AYTQGS--YSLRCNTGDSVCTTKSTRVGGTNSMGE 401
 267 WPMOVSIAHQ---NVAVCGSITTPEMIVTAACVE--KPLNPMHTATAGILROSEMEY 322
 402 WPMOVSIAHQKLAQRLLCGSLIGHQWVLAHCFGLPLQDVR--IYSGILNLSDIRK 459
 323 GAGY-OVERVISHPNTDSTKNNDAIMKLQKPLTFNDLVKPYCLPDPGMMLOPEQLCWI 381
 460 DTFPSQIKETIIHONKVSSEGNHDIKLAQAPLNTYEFQKPICLPSKGDSTITYNCWV 519
 382 SGMGAREEKGKSEVANAQVLLIETORNSRNVNLTTPAICGFLQGNVDSOGDS 441
 520 TGGFSEKGEIIONLKQVNIPLVTEECOKR--QYKTKTIRRVACVKEGKGADCKGDS 578
 442 GGPLVYSKNNIMLLIDTSGSCAKARPGVYGNVMTFTWYRMR-ADG 492
 579 GGPLVCKNKGMRVLVITSMGECARREPQVYTKAEVMDVLEKTOSSDG 630

RESULT 6
 S00845
 hepsin (EC 3.4.21.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1988 #sequence-revision 31-Dec-1988 #text-change 18-Jun-1999
 C:Accession: S00845
 R:Leytus, S.P.; Loebe, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
 Biochemistry 27, 1067-1074, 1988
 A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom
 A:Reference number: S00845; MUID:88209431
 A:Accession: S00845
 A:Molecule type: mRNA
 A:Residues: 1-417 <LEY>

A:Cross-references: EMBL:X077732; MUID:932063; PIDN:CAA30558.1; PID:932064
 C:Genetics:
 A:Gene: GDB:HPN; IMPRSS1; hepsin
 A:Cross-references: GDB:135685; OMIM:142440
 A:Map position: 19q11-19q13.2
 C:Superfamily: hepsin; trypsin homology
 C:Keywords: hydrolase; liver; serine protease; transmembrane protein
 F:23-45/Domain: transmembrane #status predicted <TM>
 F:163-400/Domain: trypsin homology <TRY>
 F:188-204-291-359-322-338-349-381/Disulfide bonds: #status predicted
 F:203-257-353/Active site: His, Asp, Ser #status predicted

Query Match 20.8%; Score 565.5; DB 1; Length 417;
 Best Local Similarity 30.7%; Pred. No. 1.2e-33;
 Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;

77 CSKTKKALCTLTGLTFLVGAALAGLTKRMGSKCSNSGIECDSSGTCIPSWCQGV 136
 12 CSRPKVA--ALTAGTLLTLAIGA-----ASMAIVA 41
 137 SHCPGGEENRCVRLGPNFIIQVSSORSKSWHPYCODDMNENTGRAACRDMGKNNFY 196
 42 VLRSDQEPFLYQVVSADARLMTVDKTEGTRLLCSSNARVAGLSCSEKGFRLALTH 101
 197 SGIYDSDSGSTSEFMKLNSTAGNVDYK-KLYHSDA-----CSSKAVYSLRCIAGV 246
 102 SELDVRFAA-----NGISGFVCYDEGRLPHTQRLLEYISVDCPRGRFLAICODCG- 154
 247 NUNSSQSRHYGGSALPGAHPWOVSLAQNVHVCSSITTEPMIVTAACVEKRLNPM 306
 155 -RRKLPVDIVIGRPTSLGRMPWOVSLRYDGAHLGGSILSDVMVLAHCPENRYLS 213
 307 HMTAFAGILROSEMEYAGYEVKYSHPN-----DSKTNDAIMKLQKPLTFNDL 360
 214 RRVYRAGAVQAQSP-HGLDQVQVAYYHGGYLPFRDPNSENDAIALVHSSPLREY 272
 361 VPEVCLPNNPMLQEPOLCWSISMGATEKGTSEYLAQVLLIETORNSRNVYNDLI 420
 273 IQPVCLPAAGALVQKICTVGMGNTQYVGOAGVLDGARVPIISNDVCADPEYNDI 332
 421 TPAMTACAGLQGNVDSQCGSGPLY-----TSKNNIMLLIGTSGSCAKARPGVYGN 476
 333 KPEKMGAGYREGIDACQDSGPFVCEDSISRTPRMRLCGIVSWGTGALAKQKGVYTK 392
 477 VAVETDWTYRMR 489
 393 VSDFERMIFQAIK 405

RESULT 7
 K002PL
 plasma kallikrein (EC 3.4.21.34) precursor - rat
 M:Alternate names: Fletcher factor; kininogenin; serum kallikrein
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Sep-1992 #sequence-revision 30-Sep-1992 #text-change 18-Jun-1999
 C:Accession: A33320; S06851; I53041; S06852
 R:Beaubien, G.; Rosinski-Chupin, I.; Matted, M.G.; Mdkay, M.; Chretien, M.; Seldah.
 Biochemistry 30, 1628-1635, 1991
 A:Title: Gene structure and chromosomal localization of plasma kallikrein.
 A:Reference number: A33180; MUID:91129236
 A:Accession: A33180
 A:Molecule type: DNA
 A:Residues: 1-638 <BEA>
 A:Cross-references: GI:J05315
 A:Note: the authors translated the codon GAG for residue 81 as Gln
 R:Seldah, N.G.; Laddenheim, R.; Mdkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; La
 DNA 8, 563-574, 1989
 A:Title: The cDNA structure of rat plasma kallikrein.
 A:Reference number: A33320; MUID:90091743
 A:Accession: A33320
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA

C:Accession: A38514; S48202; S48203
 R:Degeen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
 Genomics 8, 49-61, 1990
 A:Title: Characterization of the CDNA coding for mouse plasminogen and localization of t
 A:Reference number: A38514; MUID:91184812
 A:Accession: A38514
 A:Molecule type: mRNA
 A:Residues: 1-812 <DEG>
 A:Cross-references: GB:J04766; NID:J200402; PIDN:AAA50168.1; PID:J200403
 R:Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
 Eur. J. Biochem. 224, 863-871, 1994
 A:Title: Characterization of the murine plasma fibrinolytic system.
 A:Reference number: S48202; MUID:95010076
 A:Accession: S48202
 A:Molecule type: Protein
 A:Residues: 20-25 <LIJ>
 A:Accession: S48203
 A:Molecule type: Protein
 A:Residues: 22-27 <LIJ>
 C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many of
 C:Comment: Plasminogen is converted into plasmin by plasminogen activators, both plasmin
 immediately after dissociation from the clot. In the presence of the inhibitor, the active
 C:Comment: Streptolysin 1 (see PIR:KCMS1) acts on plasminogen to produce angiotatin. It
 etid in treating solid tumors.
 C:Function:
 A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
 n the walls of the graafian follicle; also activates the urokinase-type plasminogen act
 A:Pathway: fibrinolysis
 C:Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology
 C:Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd
 F:1-96/Domain: plasminogen-related protein precursor homology <ELPH>
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-812/Product: plasminogen #status predicted <PRO>
 F:20-96/Domain: activation peptide #status predicted <APT>
 F:79-466/Product: angiotatin #status predicted <AST>
 F:79-581,582-812/Product: plasmin #status predicted <AMT>
 F:79-581/Domain: chain A #status predicted <ACH>
 F:103-181/Domain: kringe homology <KR1>
 F:185-262/Domain: kringe homology <KR2>
 F:275-352/Domain: kringe homology <KR3>
 F:377-454/Domain: kringe homology <KR4>
 F:481-560/Domain: kringe homology <KR5>
 F:582-812/Domain: chain B #status predicted <BCH>
 F:582-805/Domain: trypsin homology <TRY>
 F:49-73,53-61,103-124,164,152-176,185-262,188-316,206-245,234-257,275-352,296-335,32
 bonds: #status predicted
 F:78-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted
 F:136,308/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted
 F:581-582/Cleavage site: Thr-Val (plasminogen activator) #status experimental
 F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 518.5; DB 1; Length 812;
 Best Local Similarity 32.7%; Pred. No. 6.8e-30;
 Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

QY 106 WKFMGSK-CNSGTECDSSGTCI-----NPSMWCQGVSHC--PGEGEDKRCVRLGPN 155
 DB 446 WEVCNLRKSETG-----GSVELLPVSOEPSPSSSETDCMKGNGKTAVAAG 499
 QY 156 FILQVYSSQKSMHPC-----QDDMNNENYGRAACRDMG-----YKNNFYSSQGIIVDSG 205
 DB 500 TPOGMAAOEPHRHSITPTQTNPRADLEKNYCRNPDDVDVNGPMXYTN-----547
 QY 206 STSFMKINTSAGNVDIKKLYHSDAGSSKAVSIRCTACGVNLSSNQ--SRITGESAL 263
 DB 548 -----PRKLY--DYCDIPLCASASFECKGPQVEPKKCPGRVYGVGVAN 589
 QY 264 PGAMPQVSLHVQ--NVHVGGSITTEPMIVTAAHCEKPLNNMHHMTAFGILROSEMF 321
 DB 590 PHSMPOWISLRTRETGQHFCGGTLIAPEWVLTAAHCEKS--SRPEFYKVIIGAHHE---644

QY 322 YGAGTQVE-----KVISHPNYDSKTKNNIDALMKLQPLFNFNDLYKVCPLPNCMLQPE 376
 DB 645 YIRLDVQVEISVAKLIEPN-----NRDIALKLSHPATITDVIPACLPSPMYVADR 698
 QY 377 QLCWISGMGATE---EKGNSEVLANAKVLLIEFGONRSRVYNDLTPMIGCGFLOGN 433
 DB 699 TICITGMGEFTGGGGRLEK---AOLPYTENKVNREYVNNRKSTELCAGQLAG 754
 QY 434 VDSQCGDSGGLVTSKNNIMWLIDTSMGSCAKAPGVGYGVNMFVDMYRMR 489
 DB 755 VDSQCGDSGGLVTFEKKDKYILQGVTSMLGACAPKPKPGYVVRVSRVDMIERMR 810

RESULT 12
 S11674
 acrosin (EC 3.4.21.10) precursor - human
 C:Species: Homo sapiens (hmn)
 C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 16-Jun-2000
 C:Accession: S11674; S23499; S12063; A61022; S03330
 R:Keime, S.; Adam, I.M.; Engel, W.
 Eur. J. Biochem. 190, 195-200, 1990
 A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gen
 A:Reference number: S11674; MUID:90306003
 A:Accession: S11674
 A:Molecule type: DNA
 A:Residues: 1-421 <KEI>
 A:Cross-references: EMBL:X54017; NID:935582; PIDN:CAA37964.1; PID:G1216165
 A:Note: the authors translated the codon AGG for residue 64 as Thr and CTr for residu
 R:Vazquez-Levin, M.H.; Reyes, J.; Gordon, J.W.
 Eur. J. Biochem. 207, 23-26, 1992
 A:Title: Molecular cloning, sequencing and restriction mapping of the genomic sequenc
 A:Reference number: S23499; MUID:92331659
 A:Accession: S23499
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-421 <VAZ>
 A:Cross-references: EMBL:M77378
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
 R:Keime, S.
 submitted to the EMBL Data Library, December 1989
 A:Reference number: S12063
 A:Accession: S12063
 A:Molecule type: DNA
 A:Residues: 1-225, 'R', 227-421 <KEI2>
 A:Cross-references: EMBL:X54017
 R:Adam, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
 Hum. Genet. 84, 125-128, 1990
 A:Title: Molecular cloning of human proacrosin cDNA.
 A:Reference number: A61022; MUID:90128988
 A:Accession: A61022
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
 R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
 FEBS Lett. 244, 296-300, 1989
 A:Title: Primary structure of human proacrosin deduced from its cDNA sequence.
 A:Reference number: S03330; MUID:89153568
 A:Accession: S03330
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-110, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
 A:Cross-references: EMBL:Y00970; NID:928325; PIDN:CAA68784.1; PID:928326
 C:Genetics:
 A:Gene: GDB:ACR
 A:Cross-references: GDB:119645; OMIM:102480
 A:Map position: 22q13-22qter
 A:Introns: 26/2; 94/2; 189/2; 237/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-421/Product: acrosin #status predicted <AMT>
 F:20-44/Product: acrosin light chain #status predicted <LCH>
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>

RESULT 14
US-08-451-932-1

Sequence 1, Application US/08451932

Patent No. 5733876

GENERAL INFORMATION:

APPLICANT: Folkman, Judah

APPLICANT: O'Reilly, Michael

TITLE OF INVENTION: Method of Treating an Angiogenic

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50

COMPUTER: Macintosh

OPERATING SYSTEM: 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/451,932

FILING DATE: 05/26/95

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/248,629

FILING DATE: 04/26/94

ATTORNEY/AGENT INFORMATION:

NAME: Larry W. Stults, Ph.D.

REGISTRATION NUMBER: 34,025

REFERENCE/DOCKET NUMBER: 05213-0123

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 812

TYPE: amino acid

TOPOLOGY: linear

US-08-451-932-1

Query Match

Best Local Similarity 19.1%; Score 518.5; DB 1; Length 812;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

DB 106 WKPMGSK-CNSGIEDSSGTCI-----NPSNMGDVSHC--PGGEDENRCVRLXGPN 155

DB 446 WEYCNLKRCEGT-----GSVELPTVSGPSPGSDSETDCMYGNGKDYRGKTAATAAG 499

DB 156 FILOVYSSQKSHNPPV-----QDDMNENYGRAACRDG-----YKNFFSSOGIVDDSG 205

DB 500 TPQGNMAAOPRHHSITPTQTNPRADLEKYNCRPDGVDVGPWCYTNN-----547

DB 206 STSFMKLNTSAGVNDYKLLYHSDACSSKAVSLRCLACGVNLSSRQ--SRIVGESAL 263

DB 548 -----PRKLY--DYCDIPLCASASSFECKPQVEPRKCGRYVGVGVAN 589

DB 264 PGAMPQVSLHVG--NVHVGGSITPEWITVTAHCEKPLNPMHTAPAGILRSFME 321

DB 590 PHSMPQWISLRTFTGQHFGGLIAPENVLTAAHCLKS--SREFYKVIILGAHEE----644

DB 322 YGAGYQVE-----KVISHPNYDSKTKKNDIALKLOKPLTFNDLVKVCILPFGMALQPE 376

DB 645 YIRGLDVOELSAVKLILERN-----NRDIALKLSPATITDKVLPACLPSPNMYVADR 698

DB 377 QLCWISGWGATE---EKGKTSEVLNAAKVLIIETQRCSNRYVYDNLITPAMICAGFLQGN 433

DB 699 TICYITMGCTGCTGFGGRUKE-----AQLPVIEKNKYNRREYELNNRKSKSELCAQLAGG 754

DB 434 VDSGQDSGGPLVTSKNNIWLIGDTSWGSCKAKAYRPGVYGVNMFETDWTYRQMR 489

DB 755 VDSGQDSGGPLVTSKNNIWLIGDTSWGSCKAKAYRPGVYGVNMFETDWTYRQMR 810

RESULT 15
US-08-452-260-1

Sequence 1, Application US/08452260

Patent No. 5776704

GENERAL INFORMATION:

APPLICANT: Folkman, Judah

APPLICANT: O'Reilly, Michael

TITLE OF INVENTION: Method of Diagnosing an Angiogenic

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50

COMPUTER: Macintosh

OPERATING SYSTEM: 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,260

FILING DATE: 05/26/95

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/248,629

FILING DATE: 04/26/94

ATTORNEY/AGENT INFORMATION:

NAME: Larry W. Stults, Ph.D.

REGISTRATION NUMBER: 34,025

REFERENCE/DOCKET NUMBER: 05213-0124

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3799

TELEFAX: 404-818-3700

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 812

TYPE: amino acid

TOPOLOGY: linear

US-08-452-260-1

Query Match

Best Local Similarity 19.1%; Score 518.5; DB 1; Length 812;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

DB 106 WKPMGSK-CNSGIEDSSGTCI-----NPSNMGDVSHC--PGGEDENRCVRLXGPN 155

DB 446 WEYCNLKRCEGT-----GSVELPTVSGPSPGSDSETDCMYGNGKDYRGKTAATAAG 499

DB 156 FILOVYSSQKSHNPPV-----QDDMNENYGRAACRDG-----YKNFFSSOGIVDDSG 205

DB 500 TPQGNMAAOPRHHSITPTQTNPRADLEKYNCRPDGVDVGPWCYTNN-----547

DB 206 STSFMKLNTSAGVNDYKLLYHSDACSSKAVSLRCLACGVNLSSRQ--SRIVGESAL 263

DB 548 -----PRKLY--DYCDIPLCASASSFECKPQVEPRKCGRYVGVGVAN 589

DB 264 PGAMPQVSLHVG--NVHVGGSITPEWITVTAHCEKPLNPMHTAPAGILRSFME 321

DB 590 PHSMPQWISLRTFTGQHFGGLIAPENVLTAAHCLKS--SREFYKVIILGAHEE----644

DB 322 YGAGYQVE-----KVISHPNYDSKTKKNDIALKLOKPLTFNDLVKVCILPFGMALQPE 376

DB 645 YIRGLDVOELSAVKLILERN-----NRDIALKLSPATITDKVLPACLPSPNMYVADR 698

DB 377 QLCWISGWGATE---EKGKTSEVLNAAKVLIIETQRCSNRYVYDNLITPAMICAGFLQGN 433

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:55:30 ; Search time 11.45 Seconds
(without alignments)
1471.939 Million cell updates/sec

Title: US-09-323-597B-2

Perfect score: 2717
Sequence: 1 MALNSGSPAIQPIYENHGY.....VYGNVMTFTMIYROMRADG 492

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0-5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2696	99.2	492	1	TMS2_HUMAN
2	2117	77.9	490	1	TMS2_MOUSE
3	875.5	32.2	454	1	TMS3_HUMAN
4	660	24.3	1035	1	ENTR_BOVIN
5	647	23.8	1019	1	ENTR_HUMAN
6	638	23.5	1069	1	ENTR_MOUSE
7	634	23.3	1034	1	ENTR_PIG
8	569.5	21.0	638	1	KAL_HUMAN
9	563.5	20.8	417	1	HEPS_HUMAN
10	558.5	20.6	855	1	ST14_HUMAN
11	556	20.5	638	1	KAL_RAT
12	548.5	20.2	902	1	ST14_MOUSE
13	538	19.8	638	1	KAL_MOUSE
14	536.5	19.7	416	1	HEPS_MOUSE
15	533	19.6	625	1	FALL_HUMAN
16	529.5	19.5	416	1	HEPS_RAT
17	518.5	19.1	812	1	PLMN_MOUSE
18	501	18.4	421	1	ACRO_HUMAN
19	490.5	18.1	415	1	ACRO_PIG
20	489.5	18.0	436	1	ACRO_MOUSE
21	486.5	17.9	761	1	NETR_MOUSE
22	477	17.6	790	1	PLMN_PIG
23	475.5	17.5	431	1	ACRO_RABIT
24	472.5	17.4	810	1	PLMN_MACMU
25	470.5	17.3	437	1	ACRO_RAT
26	470.5	17.3	810	1	PLMN_BIEU
27	461	17.0	343	1	PS38_HUMAN
28	456.5	16.8	343	1	PLMN_SHEEP
29	455	16.7	812	1	PLMN_BOVIN
30	452	16.6	875	1	NETR_HUMAN
31	451	16.6	2616	1	NDL_DROME
32	450.5	16.6	267	1	TRY7_ANOGA
33	450	16.6	270	1	TRY7_MERON

ALIGNMENTS

RESULT	1	STANDARD	PRT	492 AA.
TMS2_HUMAN				
ID	TMS2_HUMAN			
AC	015393;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	TRANSMEMBRANE PROTEIN, SERINE 2 (EC 3.4.21.-).			
GN	TPRSS2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MDL:97468144; PubMed-9325052;			
RA	Pooni-Giacchino A., Chen H., Peltsch M.C., Rossier C., Antonarakis S.E.;			
RT	Cloning of the TPRS2 gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.;			
RL	Genomics 44:309-320(1997).			
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.			
CC	- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND WEAKLY IN SEVERAL OTHER TISSUES.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.			
CC	- SIMILARITY: CONTAINS 1 SRCR DOMAIN.			
CC	- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	EMBL: U75329; AAC51784.1; ..			
DR	HSSP: P00763; IDPO.			
DR	KIM: 602060;			
DR	InterPro: IPR001254; ..			
DR	InterPro: IPR001314; ..			
DR	InterPro: IPR002172; ..			
DR	Pfam: PF00057; IdL_recept_a; 1.			
DR	Pfam: PF00089; trypsin; 1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
DR	PROSITE: PS00134; TRYPsin_HIS; 1.			
DR	PROSITE: PS00135; TRYPsin_SER; 1.			
DR	PROSITE: PS01209; LDLRA_1; 1.			
DR	PROSITE: PS50068; LDLRA_2; 1.			
DR	Hydrolase; serine protease; Transmembrane; Signal-anchor.			
KW	HYDROLASE; SERINE PROTEASE; TRANSMEMBRANE; SIGNAL-ANCHOR.			
FT	DOMAIN 85			
FT	DOMAIN 84			
FT	TRANSMEM			
FT	1			
FT	84			
FT	105			
FT	106			
FT	112			
FT	150			
FT	DOMAIN			
FT	112			
FT	150			
FT	LDL-RECEPTOR CLASS A.			

XX The invention provides a new tumour suppressor gene, designated TMPRSS2.
 CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with
 CC anticancer or therapeutic properties. The polypeptides are also useful
 CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
 CC may be used for gene therapy and protein therapy. The present sequence
 CC represents the TMPRSS2 polypeptide.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.7e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNNGSPPAIGPYENHGYOPENRYPAOPTVPTVEVHPAOPYSPPOVAPRYLTOA 60
 DB 1 MALNNGSPPAIGPYENHGYOPENRYPAOPTVPTVEVHPAOPYSPPOVAPRYLTOA 60
 QY 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMLKFMGSKCSNSGIEC 120
 DB 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMLKFMGSKCSNSGIEC 120
 QY 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNPFLQVYSSQKSWHPVCDDMNENT 180
 DB 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNPFLQVYSSQKSWHPVCDDMNENT 180
 QY 181 GRACRDMGKKNKNNYSOGIVDSDGTSFPMKLNISAGNVDIYKRLYSDACSRAVYSLR 240
 DB 181 GRACRDMGKKNKNNYSOGIVDSDGTSFPMKLNISAGNVDIYKRLYSDACSRAVYSLR 240
 QY 241 CIACGVNLNNSROSRIYGGESALPGAMPQVSLHVNQVHVCSSIIITPEWIVTAHCYER 300
 DB 241 CIACGVNLNNSROSRIYGGESALPGAMPQVSLHVNQVHVCSSIIITPEWIVTAHCYER 300
 QY 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHSPNTDSKTKNDIALMKIQKPLTFNDL 360
 DB 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHSPNTDSKTKNDIALMKIQKPLTFNDL 360
 QY 361 VKPVCILNPGMLOPEOLCWTSGMGATEEKGKTSSEVINAADVLLIETQRCNSRYVDNLI 420
 DB 361 VKPVCILNPGMLOPEOLCWTSGMGATEEKGKTSSEVINAADVLLIETQRCNSRYVDNLI 420
 QY 421 TPAMICAGFLOGNVDSCQDSCGGLPVTSKNNIMWLIGDTSMGSCCAKAYRPGVGNMVF 480
 DB 421 TPAMICAGFLOGNVDSCQDSCGGLPVTSKNNIMWLIGDTSMGSCCAKAYRPGVGNMVF 480
 QY 481 TDWYIRQMRADG 492
 DB 481 TDWYIRQMRADG 492

RESULT 2

Y44406 Y44406 standard; Protein; 492 AA.

XX Y44406;
 XX 22-MAR-2000 (first entry)
 XX Human 20P1F12-GTC2 protein.
 XX DE
 XX 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
 XX transmembrane protein; colon; prostate; prostate tumour;
 XX Homo sapiens.
 XX NC099629.42-A2.
 XX 09-DEC-1999.

PF 01-JUN-1999; 99MO-US12253.
 XX 01-JUN-1998; 98US-0087598.
 PR 29-JUN-1998; 98US-0091474.
 PR 14-APR-1999; 99US-0129521.

XX (UROC-) UROGENESIS INC.
 PA (AFAR/) AFAR D E.
 PA (HUBE/) HUBERT R S.
 PA (LEON/) LEONG K.
 PA (RAIT/) RAITANO A B.
 PA (SAFE/) SAFEFAN D C.

PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;

XX WPI; 2000-116363/10.

DR N-PSDB; 229636.

PT Novel cell surface antigen useful to treat colon and prostate cancer -

PS Claim 1, Fig 1; 58pp; English.

CC The present sequence is the 20P1F12 protein (also known as the TMPRSS2
 CC protein) which is a prostate-specific, androgen-regulated, cell surface
 CC serine protease. It is a glycosylated type II transmembrane protein with
 CC an extracellular C-terminal serine protease domain, a scavenger receptor
 CC cysteine-rich domain, an LDL receptor class A domain and a predicted
 CC transmembrane domain. Host cells can be transformed to produce this
 CC protein, using vector containing 20P1F12/TMPRSS2 gene (also designated
 CC 20P1F12-GTC1, as deposited with ATCC accession number 207097).
 CC Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for
 CC prostate and colon cancers, to image prostate cancer cells and
 CC prostate tumours, to identify ligands and cellular constituents that
 CC bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.

SO Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.7e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNNGSPPAIGPYENHGYOPENRYPAOPTVPTVEVHPAOPYSPPOVAPRYLTOA 60
 DB 1 MALNNGSPPAIGPYENHGYOPENRYPAOPTVPTVEVHPAOPYSPPOVAPRYLTOA 60
 QY 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMLKFMGSKCSNSGIEC 120
 DB 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMLKFMGSKCSNSGIEC 120
 QY 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNPFLQVYSSQKSWHPVCDDMNENT 180
 DB 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNPFLQVYSSQKSWHPVCDDMNENT 180
 QY 181 GRACRDMGKKNKNNYSOGIVDSDGTSFPMKLNISAGNVDIYKRLYSDACSRAVYSLR 240
 DB 181 GRACRDMGKKNKNNYSOGIVDSDGTSFPMKLNISAGNVDIYKRLYSDACSRAVYSLR 240
 QY 241 CIACGVNLNNSROSRIYGGESALPGAMPQVSLHVNQVHVCSSIIITPEWIVTAHCYER 300
 DB 241 CIACGVNLNNSROSRIYGGESALPGAMPQVSLHVNQVHVCSSIIITPEWIVTAHCYER 300
 QY 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHSPNTDSKTKNDIALMKIQKPLTFNDL 360
 DB 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHSPNTDSKTKNDIALMKIQKPLTFNDL 360
 QY 361 VKPVCILNPGMLOPEOLCWTSGMGATEEKGKTSSEVINAADVLLIETQRCNSRYVDNLI 420
 DB 361 VKPVCILNPGMLOPEOLCWTSGMGATEEKGKTSSEVINAADVLLIETQRCNSRYVDNLI 420
 QY 421 TPAMICAGFLOGNVDSCQDSCGGLPVTSKNNIMWLIGDTSMGSCCAKAYRPGVGNMVF 480
 DB 421 TPAMICAGFLOGNVDSCQDSCGGLPVTSKNNIMWLIGDTSMGSCCAKAYRPGVGNMVF 480

QY	1	MAINSGSPAIGPYENHNGYQPENPPYPAQPTVYVYEVHAPQAYPSPVQYAPRYLTOA	60
Db	1	MAINSGSPGIGPCYENHNGYQSEHICPPRPVAPNPNYNYLPQAYPSPVQYAPRYLTOA	60
QY	61	SNPVCYQPPSPSGTCTVCTSKTKALCLITLTLGTFYGAALAAAGLMLKFMGSKNSGTC	120
Db	61	STYIHHPKS - SCAPCTSKSKSLCLALALGTVLGAALVAAVLMLRFWDSNCSISEMC	119
QY	121	DSGCTCINPSCMGCGVHCHGCGEDENRCVLYGNFLLQVYSQKSMAPYCODDNNRY	180
Db	120	GSSGTCISSLSIMCDGVAHCHNGEDENRCVLYGQSTILQVYSQKRAMYPYCODDSEY	179
QY	181	GRACRDMGYKNNFYSSQIGYVDDSGSTFENKLTMSAGNDIYKYLHSDAGSSKAVYSLR	240
Db	180	GRACRDMGYKNNFYSSQIGYVDDSGSTFENKLTMSAGNDIYKYLHSDAGSSKAVYSLR	239
QY	241	CICACVNLNSRQSRITYGGSALPGAMPQVSLHYQNVHVCSSITTPMTYTAACHYK	300
Db	240	CIEGGR - SVKRSRIYTGGLNASGDMQVSLHYQNVHVCSSITTPMTYTAACHYK	298
QY	301	PLNNPMHTAFAGILRQSFMYAGVQVEVISHPNVDSKTRKNDIALKLPLEFNDL	360
Db	299	PLSGPRWTAFAGILRQSLMFYGRQVEVISHPNVDSKTRKNDIALKLPLEFNDL	358
QY	361	VKPCPLPNPQMLQPEOLCIVSGMGATEEKGKSEVYLMNAKVLLETQRCNSRYVDNL	420
Db	359	VKPCPLPNPQMLQPEOLCIVSGMGATEEKGKSDVYLMNAKVLLETQRCNSRYVDNL	418
QY	421	TPAMICAGFLQGVNDSCGDSGGPELYTSKNNIMWLIGDTSWGSCKKATRPVYGVNYF	480
Db	419	TPAMICAGFLQGVNDSCGDSGGPELYTSKNNIMWLIGDTSWGSCKKATRPVYGVNYF	478
QY	481	TDWYIRQMRAD 491	
Db	479	TDWYIRQMRAD 489	

RESULT 3

TMS3_HUMAN

AC P57727

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE TRANSMEMBRANE PROTEASE, SERINE 3 (PC 3.4.21.-) (SERINE PROTEASE TAG-12) (TUMOR ASSOCIATED DIFFERENTIALY-EXPRESSED GENE-12 PROTEIN).

GN TMPSR3 OR TAGD12 OR ECHOS1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

PN [1]

PP SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).

RC TISSUE-Ovarian carcinoma;

RC MEDLINE=20521358; PubMed=11068177;

RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N., Wang L., Pamley T.H., O'Brien T.J.;

RT "Ovarian tumor cells express a novel multi-domain cell surface serine protease."

RL Biochim. Biophys. Acta 1502:337-350(2000).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D).

RK PubMed=11137993;

RA Scott L.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Christ R., Guiponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S., Yonoue F., Mehdi S.O., Redhakrishna U., Papasavvas M.P., Gehrig C., Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B., Antonarakis S.E.;

RT "Insertion of beta-satellite like repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal

Best Local Similarity 78.4%: Pred. No. 1,7e+150:

Matches 385; Conservative 41; Mismatches 63; Indels 2; Gaps

[illegible]

CC prostrate neoplastic condition in an individual. Inhibitors of the
 CC proteins are useful for treating or preventing the progression of a
 CC prostrate neoplastic condition.

SO Sequence 492 AA;

Query Match 99.2%; Score 2696; DB 21; Length 492;
 Best Local Similarity 98.8%; Pred. No. 1.2e-201;
 Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAINSGSPPAIGRYENHGYOPENPYPAQPTVYEVHPAQTYSPPVQYAPRVTQA 60
 DB 1 mainsgspaiipyenhygpenypaqptvtyevhpqypqyaprvltqa 60
 QY 61 SNVYCTQPSPGTCTSTKTKALCTITLIGFLVGAALAAAGLTKFMSKCSNSGIC 120
 DB 61 snpyvctqpspgtctstktkalcitltlgtflvgaalaagllwfkmskcsngilec 120
 QY 121 DSSGTCINPSNMGDGVSHCPGEGEDENRCVRLYGPNTLOYSSQKSMHPVCODDNENY 180
 DB 121 dssgtcinpsnmgdgvshcpggedenrcvrlgpnflloyssqkshpvcqddneny 180
 QY 181 GRACRDMGKKNFYSSOGIVDDSGSTSEKMLTNSAGNVDTYKKLYHSDACSSKAVYSLR 240
 DB 181 graacrdmgyknfyssqglvddsgstslmkltnsagnvdiylklyhsdacsakavyslr 240
 QY 241 CIACGVNLSSRSRIYVGSALPGAMPQVSLHYQNVHVCGSITPEIYTAHCVER 300
 DB 241 cllcgvnlssrsgriyvgasalpgampqvslyhqvnhvcgsiltpeiytaahcver 300
 QY 301 PLNNPMTAFAGILRQSEMFYAGYOVERKVISHPYDSKTKNNDAIMKLQRLTFNDL 360
 DB 301 plnnpwmtafagilrqsemfyagyyvkvishpnydsktkndalmlkrlqrltfndl 360
 QY 361 VRVYCLPNEGMLQPEQLCWSGATGEEKTSEVLNAKVLLIETQCNSTRYVDNL 420
 DB 361 vrvyclpnegmlqpeqlcwsगतеектсевлнаквиллетқснстрывднл 420
 QY 421 TPAMICAGFLQGNVDSGCGSGLPVTYSKNNIMWLIGDTSMGSCAKARPGYGNVWF 480
 DB 421 tpmicagflqgnvdsqgsgslpvtysknnimwljgdtswgscakayrpgygnvmwf 480
 QY 481 TDWITRQMRADG 492
 DB 481 tdwlytrqmkang 492

RESULT 5
 ID Y57280 standard; Protein: 492 AA.
 AC Y57280;
 XX
 DT 06-JUN-2000 (first entry)
 DE Ovr115 homolog protein.
 XX
 KM CSG: cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
 XX endometrial; uterine; lung; cytotoxic.
 OS Homo sapiens.
 XX
 PN WO200012758-A1.
 PD 09-MAR-2000.
 PF 01-SEP-1999; 99MO-US19655.
 PR 02-SEP-1998; 98US-0098880.
 PA (DIAD-) DIADEXUS LLC.
 XX

PI Salceda S, Sun Y, Recipon H, Caferkey R;

DR WPI, 2000-256657/22.
 DR N-PSDB; Z90478.

PT Diagnosing, staging, monitoring, imaging and treating cancer especially
 PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
 PT involves measuring cancer specific gene levels in cells and body fluids

PS Disclosure; Page 52-54; 58pp; English.

CC The invention relates to detecting, diagnosing metastasis and staging
 CC cancer by measuring levels of cancer specific genes (CSG) in cells,
 CC tissues or body fluids. Their remission and progression, decreases and
 CC increases in CSG levels, is also monitored, by periodic sample analysis.
 CC The methods are useful for detecting cancers, especially gynecologic
 CC cancers which include ovarian, breast, endometrial and uterine cancer
 CC and lung cancer. Antibodies against the CSGs labeled with paramagnetic
 CC ions or a radioisotope is useful for imaging cancer and when conjugated
 CC with a cytotoxic agent are useful for treating cancer. The present
 CC sequence represents a Ovr115 homolog protein, that can be used
 CC for the detection of the various cancers.

SO Sequence 492 AA;

Query Match 99.2%; Score 2696; DB 21; Length 492;
 Best Local Similarity 98.8%; Pred. No. 1.2e-201;
 Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAINSGSPPAIGRYENHGYOPENPYPAQPTVYEVHPAQTYSPPVQYAPRVTQA 60
 DB 1 mainsgspaiipyenhygpenypaqptvtyevhpqypqyaprvltqa 60
 QY 61 SNVYCTQPSPGTCTSTKTKALCTITLIGFLVGAALAAAGLTKFMSKCSNSGIC 120
 DB 61 snpyvctqpspgtctstktkalcitltlgtflvgaalaagllwfkmskcsngilec 120
 QY 121 DSSGTCINPSNMGDGVSHCPGEGEDENRCVRLYGPNTLOYSSQKSMHPVCODDNENY 180
 DB 121 dssgtcinpsnmgdgvshcpggedenrcvrlgpnflloyssqkshpvcqddneny 180
 QY 181 GRACRDMGKKNFYSSOGIVDDSGSTSEKMLTNSAGNVDTYKKLYHSDACSSKAVYSLR 240
 DB 181 graacrdmgyknfyssqglvddsgstslmkltnsagnvdiylklyhsdacsakavyslr 240
 QY 241 CIACGVNLSSRSRIYVGSALPGAMPQVSLHYQNVHVCGSITPEIYTAHCVER 300
 DB 241 cllcgvnlssrsgriyvgasalpgampqvslyhqvnhvcgsiltpeiytaahcver 300
 QY 301 PLNNPMTAFAGILRQSEMFYAGYOVERKVISHPYDSKTKNNDAIMKLQRLTFNDL 360
 DB 301 plnnpwmtafagilrqsemfyagyyvkvishpnydsktkndalmlkrlqrltfndl 360
 QY 361 VRVYCLPNEGMLQPEQLCWSGATGEEKTSEVLNAKVLLIETQCNSTRYVDNL 420
 DB 361 vrvyclpnegmlqpeqlcwsगतеектсевлнаквиллетқснстрывднл 420
 QY 421 TPAMICAGFLQGNVDSGCGSGLPVTYSKNNIMWLIGDTSMGSCAKARPGYGNVWF 480
 DB 421 tpmicagflqgnvdsqgsgslpvtysknnimwljgdtswgscakayrpgygnvmwf 480
 QY 481 TDWITRQMRADG 492
 DB 481 tdwlytrqmkang 492

RESULT 6
 ID Y81492 standard; Protein: 283 AA.
 AC Y81492;
 XX

```

FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 842 BY SIMILARITY.
FT DISULFID 926 993 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 983 1011 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 166 192 MISSING (IN SHORT ISOFORM).
FT CONFLICT 808 808 R -> Y (IN REF. 3).
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

```

Query Match 24.38; Score 660; DB 1; Length 1035;

Best Local Similarity 33.98; Pred. No. 1.2e-41; Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

```

OY 65 VCTPKSPGVTCTSKTKALCTLT-----GTFVGAALAGLIMKMGSKCSNSG 117
DB 611 VYTGPGVNVDFSTNTMTVLFITDNMLAKGFKANFTTGGLG-----IPPCKEKN 663
OY 118 IECDSSTGTCINPSCMGVSHCGEGEDENRCVLYG-----PNFLOYSSQKRSWH 169
DB 664 FQC-KDGEICPLVNLCDGFPKCKDGSDEAHCHVRLFNCTDSSGLVQFIQ-----SIWH 716
OY 170 PVCDMDNENGRACADMGKNNFYSSGGLVDSGSIKFKMTSAGNDVIYKLVHSD 229
DB 717 VACENMTTQISDVCCLLGCTG--NSSVPTFGGPPYVNLTAPEGSLI---LPPSQ 771
OY 230 ACSKAYAVSLRC--IAGVNLNSRSR--RIYGBSALPGAMPVOVSLHYOVHVGCGSII 286
DB 772 QCLEDSLILLOQCNKSCGKLVTOEVSFKTYGSDSREGAMPVVALYFDQOYCGASLV 831
OY 287 TPFWITAAHCVERPLNPMHMTAFAGILRSFMYGAGYOVE-----KYISHPNYSKT 341
DB 832 SRDLVGAHCVYGRNNEPSKWKAVLGLHMSNL---TSPQIETRLIDQIYINHYNNRR 888
OY 342 KNNDIAMLKLOKPLTFNDLVKPVCLPNPMMLOPEQOLCWSIGMGATEEKGSTSEVLNAK 401
DB 889 KNNDIAMHLEMKVNYDYDIPICLPENQVFPGRICISIGMGALIIQGSTADVLQESAD 948
OY 402 VLLLETRCNSRYYYDNLIPAMICAGELQGNVDSGDSGGLVTSKNNNIMWLIGDSW 461
DB 949 VPLLSNECCOQMEYRN--ITENMYCAGYEAAGVDSGDSGGLMCCNNMMLLAGVTSF 1007
OY 462 GSGCAKAYRPGVYGNVAFYTDWI 484
DB 1008 GYOCALPRRPGVYARVPRFTEMI 1030

```

RESULT 5

ENTR_HUMAN STANDARD; PRT; 1019 AA.

AC P98073; DT 01-FEB-1996 (Rel. 33, Created) DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Duodenum;
RX MEDLINE=95234679; PubMed-7718557;
RA Kitamoto Y., Velle R.A., Donis-Keller H., Sadler J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinasase,
RT the proteolytic activator of trypsinogen.";
RL Biochemistry 34:4562-4568(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Holzinger A., Buck C., Maier E.M., Meyerhofer P.U., Roscher A.A.,
RA Sadler J.E., Hadorn H.B.;
RT "Genomic organization of the human enteropeptidase.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed-10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Soeda E.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Onki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Sasaki T., Shibuya K., Kawasaki S.,
RA Shintani A., Sasaki T., Nagamine K., Matsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Notides G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Degen E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lebrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE-Duodenum;
RX MEDLINE=94329561; PubMed-8052624;
RA Kitamoto Y., Yuan X., Wu O., McCourt D.W., Sadler J.E.;
RT "Enterokinasase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS--1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLY).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 NAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


NCBI_TaxID-10090;
 (1)
 SEQUENCE FROM N.A.
 STRAIN-C57BL/6; TISSUE-Duodenum;
 MEDLINE-98147142; Pubmed-9486188;
 YUAN X., ZHENG X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.;
 Structure of murine enterokinase (enteropeptidase) and expression in
 small intestine during development.";
 Am. J. Physiol. 274:G342-G349(1998).
 - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 PROTEOLYTIC PHENOMENES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PHENOMENES INCLUDING CHYMOTRYPSINOGEN,
 PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 TRYPSINOGEN.
 - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 SIMILARITY).
 - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 - SIMILARITY: CONTAINS 2 CUB DOMAINS.
 - SIMILARITY: CONTAINS 1 SCRC DOMAIN.
 - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: U73378; AAB37317.1; -
 HSSP: P00763; IDPO.
 MDG: MG1.1197533; Prss7.
 InterPro: IPR000082; -
 InterPro: IPR000859; -
 InterPro: IPR000989; -
 InterPro: IPR001190; -
 InterPro: IPR001254; -
 InterPro: IPR001314; -
 InterPro: IPR002172; -
 Pfam: PF00431; CUB; 2.
 Pfam: PF00629; MAM; 1.
 Pfam: PF01390; SEA; 1.
 Pfam: PF00530; SCRC; 1.
 Pfam: PF00057; LDL-recept.a; 2.
 Pfam: PF00089; LTPsin; 1.
 PRINTS: PR00722; CHYMOTRYPSIN.
 PROSITE: PS00134; TRYPSIN_HIS; 1.
 PROSITE: PS00135; TRYPSIN_SER; 1.
 PROSITE: PS0180; CUB; 2.
 PROSITE: PS00740; MAM.2; 1.
 PROSITE: PS01209; LDLRA.1; 1.
 PROSITE: PS01209; LDLRA.2; 2.
 Signal-anchor: Glycoprotein; Myristate; Hydroxylase;
 Serine protease; Zymogen; Transmembrane; Repeat.
 KW
 CHAIN 1 828 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
 TRANSSEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
 DOMAIN 270 379 CUB.
 DOMAIN 387 549 MAM.
 DOMAIN 569 679 CUB.
 DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
 DOMAIN 723 816 SCRC.

FT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	229	242	BY SIMILARITY.
FT	DISULFID	236	255	BY SIMILARITY.
FT	DISULFID	249	266	BY SIMILARITY.
FT	DISULFID	688	700	BY SIMILARITY.
FT	DISULFID	695	713	BY SIMILARITY.
FT	DISULFID	707	722	BY SIMILARITY.
FT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	859	875	BY SIMILARITY.
FT	DISULFID	959	1027	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1017	1045	BY SIMILARITY.
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	197	197	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	212	212	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	433	433	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	675	675	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	727	727	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	751	751	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	791	791	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	936	936	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	999	999	N-LINKED (GLCNAC. .) (POTENTIAL).
SO	SEQUENCE	1069 AA;	118735 MW;	E62549E463743C3D CMC64;

Query Match 23.5%; Score 638; DB 1; Length 1069;
 Best Local Similarity 30.5%; Pred. No. 5,3e-40;
 Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY	18 HGYPENRYPAQPTVPT	-----YEV-HPAQYPPVPOYAPAVLQAS-----	61
DB	544 NGICQSPFP-EPFLVPTPPPELPTDCGPFLEPSTFSS--PNPDPKVPQASCIWN	600	
QY	62 -----NPV-----YCTQPKSPSGVCSKTKAL	85	
DB	601 LNAQGRKNIOLHQEFLEINIDVEVRDGEEDSLLAAYTGP-GVAKDLSTTNMTV	639	
QY	86 CITLTGFLVYG--ALAAGLMKFG--SKCSNGIECDSSGTCINPNSMGVSHCPG	141	
DB	660 IFTTMMETRRKGFKAANTSQY--YLGIPRCQDDEQC-KDGNCTPLGLNCSYPRCRD	715	
QY	142 GEDENRCVRLY---GNFLLQVYSQKSMHPYQDDMMENYGRAACRDMGKNNFYSS	197	
DB	716 GSDASCVRFLNGTRSNNGVLQF--NIHSIMHICAEIMMTQIISNEYCHLGLCS--ANS	771	
QY	198 QGITDDSGSTFEMKLNISAGVNDIYKLIHSDACSSAAVSLRC--IAGCVNLSSNOS-	254	
DB	772 SMPISSTGGEFVRVNDAPNGSLI--LTPSLQCSQSLILLCNHRKSCGKKYQVSP	828	
QY	255 RIVGESALPGAMPQVSLAHQNVH---VCGGSITPEMIVTAAHCVERPPLNPMHTA	310	
DB	829 KIVGSDAQAGAMPVVALYHRDSTRLCGASLVSDMIVSAHCAVYRNRNDPTMTA	888	
QY	311 FAGILROSFMYGAGYO--VEKVISHPNDYSKTKNNDIALMKLQKPLTFNDLVKVCPLN	368	
DB	889 VLGHMOSNLISPOVVRVAVDQIVINPHYDRKRVNDIAMHLEFKVNYDYIOLPILPE	948	
QY	369 PGAMLOEOLCISGKATE-EKKTSEVYLNAAKVLLIEORCSRVRVNDLTPAMICA	427	
DB	949 ENQIFIGRTCSIGAGWYDKINAGSTVDYKADVPILSNKCCQOQAPERYN-ITESHICA	1007	
QY	428 GFLGNDVSCGDSGGFLVYTSKNNIMWILIGDTISGSCAKAYRPGVYGNVAFPTDWY	485	
DB	1008 GYEGGIDSCGDSGGFLMCOENNRWFLVGVTSFGVOCALPNHGVYVVRYSQFLEWH	1065	

Db 772 CFEESLILLOCNHAKSCGKQVQAEVSKYGVGNDSEGANPWTVALYNGLLGASIVS 831
 QY 288 PEMIVTAHCVKELPNMHTAFAGILROSFMYGAGYQ-----VEKVISHPNYDSKTK 342
 Db 832 RDMVSAHACHVGNLLEPSKKKALILGLHMTSNL---TSPQVTLIDELIVPNHNRKK 888
 QY 343 NNDIALMLKLPFLFNDLVKFCVCLPNPGKMLQEPOLWISSWGATEEKGTSEVNLNAKV 402
 Db 889 DSDIAMHLEKRYVYTDYIDPICLPEENQVFPGRICSIAGMGVITYOGSPADLIQEAADV 948
 QY 403 LLITQRCNSYVVYDNLITPMICAGFLGVNDSCOGDSGGELTYSKNIMVLLIGDTSNG 462
 Db 949 PLTSENCCQOQMPRYN-ITENNMCAGTIEEGSIDSCGGSGPLCMLENRMVLAGVTSFG 1007
 QY 463 SGCAKAYRPGYGVNMYFTDWI 484
 Db 1008 YQCALPMPRPYARVPKFTEMI 1029
 RESULT 8
 KAL_HUMAN STANDARD; PRT; 638 AA.
 AC P03952;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KIK3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=6243359; PubMed=3521732;
 RA Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;
 RT "Human plasma prekallikrein, a zymogen to a serine protease that
 RT contains four tandem repeats.";
 RL Biochemistry 25:2410-2417(1986).
 RN [2]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=91152016; PubMed=1998666;
 RA McMullen B.A., Fujikawa K., Davie E.W.;
 RT "Location of the disulfide bonds in human plasma prekallikrein: the
 RT presence of four novel apple domains in the amino-terminal portion of
 RT the molecule.";
 RL Biochemistry 30:2050-2056(1991).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XII, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- DISEASE: DEFECTS IN KIK3 ARE THE CAUSE OF FLETCHER FACTOR
 CC DEFICIENCY, A BLOOD COAGULATION DEFECT.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13143; AAA60153.1; -
 CC DR PIR; A00921; KOHUP.

DR PIR; A37939; A37939.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.212; -.
 DR MIM; 229000; -.
 DR InterPro; IPR000177; -.
 DR InterPro; IPR001254; -.
 DR InterPro; IPR001314; -.
 DR InterPro; IPR003014; -.
 DR Pfam; PF00024; PAN; 4.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00005; APPLIEDOMAIN.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00495; APPLE; 4.
 KW Hydrolysis: Serine protease; glycoprotein; plasma; zymogen; signal;
 KW Fibrinolysis; blood coagulation; inflammatory response; liver;
 KW Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 390
 FT CHAIN 391 638
 FT DOMAIN 20 105
 FT DOMAIN 110 195
 FT DOMAIN 200 285
 FT DOMAIN 291 376
 FT DOMAIN 389 621
 FT CARBOHYD 127 127
 FT CARBOHYD 308 308
 FT CARBOHYD 396 396
 FT CARBOHYD 453 453
 FT CARBOHYD 494 494
 FT ACT_SITE 434 434
 FT ACT_SITE 483 483
 FT ACT_SITE 578 578
 FT DISULFID 21 104
 FT DISULFID 47 77
 FT DISULFID 51 57
 FT DISULFID 111 194
 FT DISULFID 137 166
 FT DISULFID 141 147
 FT DISULFID 201 284
 FT DISULFID 227 256
 FT DISULFID 231 237
 FT DISULFID 292 375
 FT DISULFID 318 347
 FT DISULFID 322 328
 FT DISULFID 340 345
 FT DISULFID 383 503
 FT DISULFID 419 435
 FT DISULFID 517 584
 FT DISULFID 548 563
 FT DISULFID 574 602
 SO SEQUENCE 638 AA; 71369 MW; E62P9C1053038PB4 CRC64;
 Query Match 21.0%; Score 569.5; DB 1; Length 638;
 Best Local Similarity 32.8%; Pred. No. 37e-35;
 Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;
 QY 55 RVLQASNPVYCTPKPSGVCYSKTRKALCTLTGTPLVGAALAGLLMKRMSG--- 111
 Db 218 RVLV--DPAFCR-----TICTYHNP--CLFTEYV-----NWKLTESGRNV 255
 QY 112 ---RCSNGIRCDG-----SG-----TCINP-----SNMCDVSHCPGGEDENRVRL 151
 Db 256 CLKLTSSGSPSSSTPDENTISGSLTCKRTLPKPGSKRIYVDV--GGEELN----- 308
 QY 152 YGNPFIQVYSSQKSWHPYVCOODWNNYGRACRDNGYKNNFYSSGQIVDSSGYS--- 208
 Db 309 --YVFKGV-----NVCQE-----TCTKM-IRCOFTYSLLPEDCKEERCKC 347
 QY 209 FMKLTNSAGNDYIKKILYHSDACSSKAVVSLRCLAGVN--LNSRSRSRTVGGESALPGA 266

DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
 DE TYPE SERINE PROTEASE 1) (MT-SP1).
 GN S714 OR PRS514 OR SNC19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=9303581; PubMed=10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RA "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
 RT protease with trypsin-like activity."
 RL J. Biol. Chem. 274:18231-18236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RA "Reverse biochemistry: Use of macromolecular protease inhibitors to
 RT dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 RN [3]
 RP CHARACTERIZATION.
 RP TISSUE=Milk;
 RX PubMed=10373425;
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
 RA "Purification and characterization of a complex containing matriptase
 RT and a Kunitz-type serine protease inhibitor from human milk."
 RL J. Biol. Chem. 274:18237-18242(1999).
 CC -1- FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
 CC OR LYS AS THE P1 SITE.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb.slb.ch/announce/>
 CC or send an email to license@lsb.slb.ch).
 CC -----
 DR EMBL: AF118224; AAD42765.2;
 DR EMBL: AF133086; AAF00109.1;
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.302;
 DR InterPro: IPR000859;
 DR InterPro: IPR001254;
 DR InterPro: IPR001314;
 DR InterPro: IPR002172;
 DR Pfam: PF00057; 1dl_recept_A; 4.
 DR Pfam: PF00089; trypsin_1.
 DR Pfam: PF00431; CUB. 2.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS. 1.
 DR PROSITE: PS00135; TRYPSIN_SER. 1.
 DR PROSITE: PS01180; CUB. 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS50068; LDLRA_2; 4.
 DR Signal-anchor: Glycoprotein; Hydrolase; Serine protease;
 KM Transmembrane: Repeat.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 214 334 CUB 1.
 FT DOMAIN 340 447 CUB 2.
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 614 851 CATALYTIC.
 FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
 Query Match 20.6%; Score 558.5; DB 1; Length 855;
 Best Local Similarity 33.9%; Pred. No. 3,4e-34;
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;
 QY 110 GSKGNSGIECDSSGTCINPMDGVSHCPGDEENRCVRLGPNFLLOYSSQKSMH 169
 DB 522 GCSCPAQTFRC-SNGKLSKSGQCNGKDDCGSDSASCPV---NVV----- 565
 QY 170 PVCDDMNENYGRACSDMGYKNNFYSSQGITVDGSGTSEKMLTSGANDIYKLYHS 229
 DB 566 -----TCT-----KHRYRLANG-----LTLKGNPECDKEDCSD 595
 QY 230 ACSKAYSLRCLACGVNLS-SQSRHYGESALPQAMPQVSLH-VQNVHVGGSIT 287
 DB 596 GSDEK-----DCDGLRFTQARVAGTDADEGEMPVSLHLDGCHIGASLIS 647
 QY 288 PEWYTAHCVCEKPE---LNNPMHTAFAGILRSFMEYAGAYO---VEKYISHPNYSK 340
 DB 648 PNLVSAHCHYIDRGRFSDPTQMTAFGLHDQS-QRSAPGVERHLKRLTSHPFNDF 706
 QY 341 TKNDIALMKLQKPLFENDLVKPYCLPFGMLQPEOLCWTISGATBEKKGISEVLNAA 400
 DB 707 TFDYDIALLELEKPAEYSSMVRPCLDPASHVFPAGKAIVTGWGHGYGTGALLILQKG 766
 QY 401 KVLIEPQRCNSRVYNDLILPAMICAGFLQGNVDSGCGSGPL--VTSNNINIMLIGD 458
 DB 767 EIRYNOTTCEN--LLPQGITPRMVCGLFSGVDSGCGSGPLSVYADGRI-FOAGV 823
 QY 459 TSMGSCAKAYRPGVYGVNVAFTDPI 484
 DB 824 VSMGDGCAQRKPKGVYTRRLPLFRDPI 849
 RESULT 11
 KAL_RAT STANDARD; PRT; 638 AA.
 ID KAL_RAT
 AC P14272;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN PK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129236; PubMed=1993180.
 RA Beaubien G., Rosinski-Chupin I., Mattei M.-G., Miley M., Chretien M.,
 RA Seidah N.G.;
 RA "Gene structure and chromosomal localization of plasma kallikrein."
 RL Biochemistry 30:1628-1635(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```

Db 594 SDGDER-----NC-DCGLR-SFTKQARVYGGTNADEGENPMQVSLHAIQGHLCGASLI 646
QY 287 TPETIVYPAHCVKPEPLNNPWH-----WPAFAGILNOSFFFTYAGVQ---VEVYISHPNDS 339
Db 647 SPDLVSAHCFQDDKFKFSYDMYMAAFGLGLDQS-KRSASGVQELAKLILTPSPND 705
QY 340 KTKNNDAIAMLKRLPFLFNFDLVKPCVCLPNPQMALQPEOLCISGSGATEERKGTSEVNA 399
Db 706 FTPEYDIALLELEISVEXSTYVRICLPDAHHVPAAGAIWVTHGMHKBEGTGALLQK 765
QY 400 AKVLLIFTORCSRNSRYVDNLITPAMICAGFLQGVNDSCQDSSGGPELVTS-KNNIWLIGD 458
Db 766 GEIVINQTCED-LMPQGITPRMCMGVFLSGVSCQDSSGGPLSAERKGRMFGV 823
QY 459 TWSGSCAKATPEVY 474
Db 824 VSMGEGCAORNKPGVY 839

RESULT 13
KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KIK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1;TaxID=10090;
RX [1]
RA Seldan N.G., Sawyer N., Hamelin J., Mion P., Beauplen G.,
RA Brachpapa L., Rochemont J., Molikay M., Christien M.;
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZWINGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M58588; AAA63393.1; .
DR PIR; A36557; KOMSPL.
DR HSSP; P00750; IRFE.
DR MEROPS; S01.212; .
DR MGD; MGI:102849; KIK3.
DR InterPro; IPR000177; .
DR InterPro; IPR001254; .
DR InterPro; IPR001314; .
DR InterPro; IPR003014; .
DR Pfam; PF00024; PAN; 4.

```

```

OY 137 SHCGGDEKNCVRLYGPNNFILOYSSQSKSMHPVCCODMNNENYGRAACRDMGKNNFYS 196
DB 41 ILLOSDEPLYQVQLSGCDRLAVLDTKTEGTWLLCSSRNARVAGLCSEMGFLRLAH 100
OY 197 SGGIVDDSGSTSEFKMLNTSA-----GNVDIYKKLYHSDA---CSSKAVVSLRCIAGVN 247
DB 101 SELDVRNAGAN-----GTSGEFCVDEGGLPLAQLRLLVISVCCDPRPRFLATQDDG-- 153
OY 248 LNSRSRIGVGSALPGAMPQVSLHVQNVHCGSGIIPPEWIVTAHCEKPLNPMH 307
DB 154 RRLKPVDRIVGQSSSLGRPMQVSLRYDGTHTLCCGSLLSGDWTLTAHCFEPERNVLSR 213
OY 308 WTPAFAGILRSFMYGAGVYKATISPNV-----DSKIKNNIDIAKMLQKPLTFNDLV 361
DB 214 WRFYAGAVARTSP-HAVQLGVAQVAYHGYLPRFDPRTIDENSNDIAVLHSSSLPLEYEI 272
OY 362 KPVCLPAPGMMLOPEOLCMTISGMATEKSTSEVLNAAKLLIETORCNSRYVDNLIT 421
DB 273 QPVCLPAPGALVDGKVCYTGKMGNTQFYGOQANVLOEAPRTISNEVCNSPDYGNQIK 332
OY 422 PAMICAGFLOGNVDSCGDSGGPLY---TSKNIMWLIGDTSWGSQCAKARPGYGVN 477
DB 333 PKMFCAGYPEGIDACGDSGSPVCEDSISGTSRMRLCGIVSGTGALARKGYVTKV 392
OY 478 WFTDWIYRQNR 489
DB 393 TDFREWIFKAIK 404

RESULT 15
FAIL_HUMAN STANDARD; PRT; 625 AA.
ID PO3951;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
DE ANTECEDENT) (PTA).
GN F11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86243360; PubMed=363155;
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor
RT with four tandem repeats that are highly homologous with plasma
RT prekallikrein."
RL Biochemistry 25:2417-2424(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=68107663; PubMed=2827746;
RA Asakai R., Davie E.W., Chung D.W.;
RT "Organization of the gene for human factor XI."
RL Biochemistry 26:7221-7228(1987).
RN [3]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152017; PubMed=1998667;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the
RT presence of tandem apple domains."
RL Biochemistry 30:2056-2060(1991).
RN [4]
RP VARIANT LEU-301.
RX MEDLINE=90046636; PubMed=2813350;
RA Asakai R., Chung D.W., Ratnoif O.D., Davie E.W.;
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi
RT Jews is a bleeding disorder that can result from three types of point
RT mutations."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).

```

```

RN [5]
RP VARIANT LEU-301.
RX MEDLINE=92190478; PubMed=1547342;
RA Metjers J.C., Davie E.W., Chung D.W.;
RT "Expression of human blood coagulation factor XI: characterization of
RT the defect in factor XI type II deficiency."
RL Blood 79:1435-1440(1992).
CC -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL
CC BONDS IN FACTOR IX TO FORM FACTOR IXA.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND. AFTER ACTIVATION
CC THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- PTM: ACTIVATED BY FACTOR XIIa (OR XII), WHICH CLEAVES EACH
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH
CC MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION
CC ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN
CC ASHKENAZI JEWS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13142; AAA52487.1; -
DR EMBL; M20218; AAA51985.1; -
DR EMBL; M18296; AAA51985.1; JOINED.
DR EMBL; M21184; AAA51985.1; JOINED.
DR EMBL; M18298; AAA51985.1; JOINED.
DR EMBL; M18298; AAA51985.1; JOINED.
DR EMBL; M18300; AAA51985.1; JOINED.
DR EMBL; M18301; AAA51985.1; JOINED.
DR EMBL; M18302; AAA51985.1; JOINED.
DR EMBL; M18303; AAA51985.1; JOINED.
DR EMBL; M18304; AAA51985.1; JOINED.
DR EMBL; M19417; AAA51985.1; JOINED.
DR EMBL; M20217; AAA51985.1; JOINED.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.213; -
DR MIM; 264800; -
DR InterPro; IPR000177; -
DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR InterPro; IPR003014; -
DR Pfam; PR000024; PAN; 4.
DR Pfam; PR000089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Hydrobase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW duplication; Signal; Disease mutation.
FT SIGNAL 1
FT CHAIN 19 387
FT CHAIN 388 625
FT DOMAIN 19 104
FT DOMAIN 109 194
FT DOMAIN 199 284
FT DOMAIN 290 375
FT DOMAIN 384 625
FT CARBOHYD 90 90
FT CARBOHYD 126 126

```


301 PLNNPMHTATAGILIRSGFMFYGAGYOEKXISHPNYSKTKNNNDIALMLOKPLTFNDL 360
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 299 PLSGPRYTATAGILIRSGFMFYGAGYOEKXISHPNYSKTKNNNDIALMLOKPLTFNDL 358
 361 VKPYCLPMPGKMLPPEOJLCWISGMCATEEKGKTSSEVLNAARVLLIETORCSRYVDNLI 420
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 359 VKPYCLPMPGKMLDLDGECWISGMCATEEKGKTSSEVLNAARVLLIETORCSRYVDNLI 418
 421 TPANICAGFLOGNDSGCGSGGLYTSKNNIMWLIGDTSNCSGCAKAVRPVYGNWYF 480
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 419 TPANICAGFLOGSDSCGSGGLYTSKNNIMWLIGDTSNCSGCAKALRPVYGDVYF 478
 481 TDWIYQMRAD 491.
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 479 TDWIYQMRAN 489
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 RESULT 4
 09NZAS PRELIMINARY; PRT; 423 AA.
 09NZAS
 AC 09NZAS:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE TYPE II MEMBRANE SERINE PROTEASE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
 RT "M-92, a novel type II membrane serine protease expressed in
 RT trachea, colon, and small intestine: identification, cloning, and
 RT chromosomal localization.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF216312; AAF31436.1; .
 KM Protease.
 SO SEQUENCE 423 AA; 46397 MM; 90792AF08AFE30 CRC64;
 Query Match 25.2%; Score 684; DB 4; Length 423;
 Best Local Similarity 36.2%; Pred. No. 5.8e-54;
 Matches 165; Conservative 64; Mismatches 167; Indels 60; Gaps 16;
 61 SNPVCCTQPKASPSGVCSTKTKKALCITLTGTFLYVGAALAGLTKFKSGKCSNGIEC 120
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 2 SNP--CANPVPD-WRPSVSGIPITIALSLASTIIVYVLLIVLDKTY-----FLC 50
 121 DSSGTCINPMNWCDSVSHGCGEDENRCVRLY--GP-----NFLQVYSQKRSWH 169
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 51 GQPLHFIRKQLODELCPLEGDEHCVKSPPEPAVAVRLSKDRSTLOYDSATGMWF 110
 170 PYCQDDNMENNGRAACRMGY--KKNFSSQ-----GIYDSSSTSMKNTSAGAND 220
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 111 SACEPNFTEALAEACRMGYSSKPTFAVEGPDQDDIVYEITENSOELRNRRNSG-- 167
 221 IYKKLHSDACSSKAVYSLRGIACVNLNNSHQRIVGESALPGAMPQVSLHYQNVHY 280
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 168 -----PCLSGSLVHLCLACSKSL---KTPRVYGGEEASVSWPQVSIQYDKQHY 215
 281 CGSITTPPEWIVTAACHCEKPLNNMHTAFAGILR-OSFMFYGAGYOEKXISHPNYSKTKNNNDIALMLOKPLTFNDL 360
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 216 CGGSITDPHWVLLTAHCRKH-TDVENKMKVRAKSGDKLSF-----PSLAIAKIIITEFNPM 270
 337 YDSKTKNNNDIALMLOKPLTFNDLYKPCPLPMPGKMLDLDGECWISGMCATEEKGKTSSEVLNAARVLLIETORCSRYVDNLI 420
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 271 Y--PKNDIALMLOKPLTFNDLYKPCPLPMPGKMLDLDGECWISGMCATEEKGKTSSEVLNAARVLLIETORCSRYVDNLI 420
 396 VLNAKVLIIETORCSRYVDNLTTPANICAGFLOGNDSGCGSGGLYTSKNNIMWL 455
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111
 Db 328 ILLQASVQVISTKRNADDAIAGEVTEKMKMGKIGTPEGVDTQCGSGGGLMLOKPLTFNDL 360
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::||||| 1111

OY	456	IGDTSWGSACAKARPGVYGNNVFETDIYKQMRAD	491
		: : :	:
Dd	387	VGIYSMGYCGGSPSTPCGYTTKVSAYLNMIVNWKAE	422
RESULT	5		
ID	09NRSA	PRELIMINARY;	PRT; 437 AA.
AC	09NRSA		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	TRANSMEMBRANE SERINE PROTEASE 3.		
GN	TMPRSS3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PANCREATIC CARCINOMA;		
RX	MEDLINE=20283276; PubMed=10825129;		
RA	Wallirpp C., Hammel S., Muller-Pillasch F., Burghardt B., Iwanura T.,		
RT	Ruthebuenger M., Leich W.M., Adler G., Gress T.M.;		
RT	"A novel transmembrane serine protease (TMPRSS3) overexpressed in		
RL	pancreatic cancer.";		
DR	Cancer Res. 60:2602-2606(2000).		
KM	EMBL: AF179224; AAF74526.1; .		
SQ	Protease. 437 AA; 48204 MW; 351B2FDA08657B12 CRC64;		
Query Match	24.9%; Score 676.5; DB 4; Length 437;		
Best Local Similarity	39.18; Pred. No. 2.9e-53;		
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;			
OY	133	CDGVSHCPGGDENRCVRLY--GP-----NEILQVYSQRKSHPVCODDNENYNG	181
		: : :	:
Dd	77	CDGEIDCPIDGEDEHCYVSPFEGPAVAVALRSLKDRSTLOVLDSATGMFMSACFDNFTAL	136
OY	182	RAACHDMKY--KNNYSSQ-----GIVDDSGSFMKLNTSGANDVIYKKLYHSDACS	232
		: : : : : : : :	:
Dd	137	ETACKOMKSISSKPTFRAYEIGPDODLDVETENSQELMKRNSSG-----PCL	184
OY	233	SKAVYSLCIAAGVALNSSROSRIYGESAPRGAMPQOVSLHYOVNVHVGGSIIITPENIV	292
		: : : : : : : : : :	:
Dd	185	SGSLVSLHLCAKGSLS---KTPRYVGGESASVDSPMVQSIOYDKQHCVGGSIIDPHVYL	241
OY	293	TAAACVEKPLANNPMHWTAFAGITLR-QSMFYGAGIYQVKYV---SHPNYDSKTKNNDIAL	348
		: : : : : : : :	:
Dd	242	TAAHGFRKH-TDVFWKWRKRAAGSDIKGSE----PSLAIVAKIIIEFPNY---PKNDIAL	293
OY	349	MKIQLPLFNLDLVKPYCLPDPNGMLQPDLQWLISGMGAITEEK-GKTSVLNAKULLLET	407
		: : : : : : : :	:
Dd	294	MKLOPLPLFFSGTVPRICLPFFDEDELTRPTPLMIIGMTFKONGKMSIIDLQASYVIDS	353
OY	408	ORCNRSRYVYDNLITPAMICAGFLQGNVDSQGDGSGPLVTSSKNNIWMILIGTSMWGSCAK	467
		: : : : : : :	:
Dd	354	TRCNADAYGGEVTEHKMGACAGIEGGVDTCCGDGSGPLMYGSDQ-WHVYGIYSMGYCGCG	412
OY	468	AIRPGVYGNNAVETDWIYKQMRAD	491
		: : :	:
Dd	413	PSTPGVYTKVASAYLNMIVNWKAE	436
RESULT	6		
ID	09Y505	PRELIMINARY;	PRT; 1042 AA.
AC	09Y505;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-Oct-2000 (TREMBLrel. 15, Last annotation update)		

```

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tazuki S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB037898; BAB03502.1;
KW Protease.
SO SEQUENCE. 855 AA; 94955 MW; 358067E6CF6CF03D CRC64;

Query Match. 20.8%; Score 564.5; DB 11; Length 855;
Best Local Similarity 33.0%; Pred. NO. 1e-42;
Matches 128; Conservative 62; Mismatches 131; Indels 67; Gaps 11.

OY 110 GSKCNSGIECDSSGTCINPMNCDGVSHCPGGEENRCVRLYGNPFIQVYSSQKSMH 169
DB 522 GCSCPAAGSFKC-SNKKCLPQSQGCKDGDGSDBASCDNVNAVASCRTKYTRCO----N 576
OY 170 PYCQDDMN-ENTGRACRDMGTYKNNFYSSQGLVDSGSISFEMKLTNSAGNDIYKKLYHS 228
DB 577 GICLKNKGNEPCGKKKDCSDGSEKNC-----DCGLNSFTK----- 611
OY 229 DACSSKAVYSLRCLACGVNLSSROSRIYGEESALPGAMPQVSLH-YONVHVCGSIT 287
DB 612 -----DARYVGGTNADEGENPQVSLHALQGHLCGSLIS 647
OY 288 PEMITVAHCVYK-----PLNNPMHTAFAGILROSFMEFYAQY--YKVYISHPNYDSK 340
DB 648 PDMVLVAHCFQDEITFKYSDHTMTATFAGLLDQS-KRASGVQEHKELRITHPSENF 706
OY 341 TKNDIALMKLQKPLTFNDLVKPVCLPNPDMKMLQBPQLCWISGCGATEKGTSEYLYNA 400
DB 707 FEDYDIALLELEKPKAYSTVYRPICLPDHTVHFPAKAIWTVGWGHTKEGTCALILIKG 766
OY 401 KVLLETQRCNSRRYYNDNITPAMICAGELQGNVNSCCDSDGSEPLVT-SKNINWMLIGPT 459
DB 767 EIRVINGTTCEE--LLPQGITPRMNCVGLSGGVSDCGSDGSSVSKGRIFQAGV 824
OY 460 SMGSCAKARPGVYGNVVFETDIYRQ 487
DB 825 SMGEGCAQRNKGVTTRIFEVARDWIKQ 852

RESULT 9
O97506 PRELIMINARY; PRT; 643 AA.
AC O97506;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NC NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Takehashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB022425; BAA37147.1;
DR HSSP; P00766; 1CHG;
DR MEROPS; S01.212; -.
DR INTERPRO; IPR00177; -.
DR INTERPRO; IPR001254; -.
DR INTERPRO; IPR001314; -.
DR INTERPRO; IPR003014; -.
DR PFAM; PF00024; PAN; 4.
DR PFAM; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.

```

Query Match 30.7%; Score 563.5; DB 6; Length 643;
Best Local Similarity 29.9%; Pred. No. 8,86-43;
Matches 154; Conservative 63; Mismatches 105; Indels 113; Gaps 19;

DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
DR PROSITE: PS00495; APPLE; 4.
SQ SEQUENCE 643 AA; 7227 MW; AFF9232E3C3CB80A CMC64;

Query 52 YAPRYLQASNPVYCTQPSGSGYVTSKTKKALCTLTGLTGLAALACLKFKNGS 111
Db 160 YATQAFNAEYRNKCNLKHSPGFTPSIK-----VLAVNESG-----FSLK 200
QY 112 KCSNSGICDSS-----GTCINPSNMGDGVSHCHGGEDENCVRLYGNFIQ 159
Db 201 PCADSEICCHNDIFQHLAFSDVDVARIAPAFV-----C-----PTICYHPNCLFF 248
QY 160 VY-----SSOR-----KSMH-----PVCODDNNENYGRAACRDM---GYKNPFYSQ 198
Db 249 TFYYNAWKIESQRNVCFLKTHSGTSPPTFOENALISGYSLLTCTQTLPEPCHSKITYEV 308
QY 199 GIVDSDGSTFEM-----KLNTSAGNYDIKKLYHSDAC--SSKAVSLR----- 240
Db 309 DFEGBELAVTFYQGANLQGETCTKTRIQCFPTLYSLHPEDCREKCKSLRISDGSPTKI 368
QY 241 -----C-----IACGVNLNLSRSRSLRYGEGSALPGAMPQVY-----LHYON 277
Db 369 TGHMRASSGYSLRLCRSDHSHACATKAN---TRVVGDTDFSLGMPQVYSLOAKTLRAON 424
QY 278 VHVCGGSLTTEEMVITAHCHVEKPLNNNMHMTAFAGILLRQSMFEGAGY-OYEKVISHPN 336
Db 425 -HLCGGSLTGHQWVLTAAHCFD-GSLDPIWMYIGGLINISLETETPEPSQVKEIITHON 482
QY 337 YDSKTKNNDIALMKLQKRLFTENDLVKPYCLPNPQMALPEOCLWISMGATEKEKSTSEY 396
Db 483 YKLESQGDIALKLETPLTNTDTPKPLCLPSRDTNTVYTCWTTGKCFTEKEKLEION 542
QY 397 LNAAKVLLIETQRCNSRRVYDNLITPAMICAGFLQNVDSGCGSDSGPLVTSKNNIMLI 456
Db 543 LQKNVPIPLVNSEQCKSY-RDHKISKQMICAGYKRGKADCKGSGGSLVCKYNGIMHLY 601
QY 457 GDTISWGGCAKARPGVYGVNAVTFEDMYLRQMRAD 491
Db 602 GTTSGEGCARREOPGVYTKVIEYMDWILEKTODD 636

RESULT 10
Q9Y5Y6 PRELIMINARY; PRT; 855 AA.
AC Q9Y5Y6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE MARRIPASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303581; PubMed=1037424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine
RL protease with trypsin-like activity";
RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.


```

Db 922 AKCRGCGDEIYCDSDS-QCIGTKHICDGIIDCPYGODERNCLRLSERNGDVGTGLEVYR 980
QY 163 SQRSMHPCVODDNNENNG-BAACRDMGKNNFYSSQIYVDSGSTSPMK-----LNT 214
Db 981 IGOQMPACVCKNMDRAVSPSAVCSILCY-----SAVNTSVTLTILTRPPLAT 1029
QY 215 SAGNVDIYK-----KLYHSDACSSKA-----VYSLRC-IAGC-VNLNRSRQS-RIVGG 259
Db 1030 VNVSTDLIMKMYAKRSTLMQEFANCKTEDYPMADLTCNSNECGRVKGRKPSRRIGG 1089
QY 260 ESALPGAMPVQSL-HVQNVHVGCGSITPEMAYTAHCY-EKPLNPNMWTAFAGTLR 316
Db 1090 TQASPGMPPFLAAILGCGPEKIFYCAGVLYSDQWVLTASHCGVNSVLDLEMTIOLGVTYR 1149
QY 317 Q-SMFGAGQVQVKVISHPNVD-SKTKNNDIALMLOKPLTFNDLVKPYCLPNDGM-L 373
Db 1150 RNSFTYSGQKVKVAVLPHPOYNMAIHNDIALFQLATRAFEHHLPLVCLPPSVNRL 1209
QY 374 QPEOLCWSGATEEKGKTSSE---VLNAKVLLETQRCNSRYVDNL-ITPAMICAGF 429
Db 1210 HPGTLCTVYIGKREDKDPKSTYEIVNEVQPIITRNQDE--WLDNLTVSEGVACGAF 1267
QY 430 LQGVNDSCQDGSGLV---TSKNNIMWLIDGTSWGSACAKAYRPGVGNVVEFTDWIYR 486
Db 1268 DDGKDAQCQDGSGLPLCLPYGKRNRFVGGIVSGWICAHPLRPLGVYANVYQVYPIQOE 1327
QY 487 QM 488
Db 1328 QI 1329

```

```

RESULT 12
Q9Y495 PRELIMINARY: PRT: 571 AA.
AC Q9Y495;
DT 01-NOV-1999 (TREMBLREL. 12, Created)
DT 01-NOV-1999 (TREMBLREL. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
DE PLATELET FACTOR XI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86256306; PubMed=9593722;
RA Hsu T.C., Shore S.K., Seshima T., Bagasra O., Walsh P.N.;
RT "Molecular cloning of platelet factor XI, an alternative splicing
RT product of the plasma factor XI gene."
J. Biol. Chem. 273:13787-13793(1998).
DR EMBL: AF045649; AAC24506.1; -
DR HSSP: P20231; IAAO.
DR INTERPRO: IPR001177; -
DR INTERPRO: IPR001254; -
DR INTERPRO: IPR001314; -
DR INTERPRO: IPR003014; -
DR PRAM: PR000024; PAN: 3.
DR PRAM: PF000089; trypsin. 1.
DR PRINTS: PR00005; APLEDOMAIN.
DR PRINTS: PR00122; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00495; APPEL: 3.
SQ SEQUENCE 571 AA; 63840 MW; FB9D65D72151755E CRC64;

```

```

Query Match 19.6%; Score 533; DB 4; Length 571;
Best Local Similarity 30.5%; Pred. No. 4.5e-40;
Matches 122; Conservative 63; Mismatches 129; Indels 86; Gaps 11;
QY 169 HPCV-----ODNNENYGRACRDMGKNNFYSSQIYVDSGSTSPMKLNTSAGNDIY 222
|||||

```

```

Db 179 HPGCLFFTFPSQEMPKRSQNRNLCLKTSEGLPSTR--IKSKRLSGFSIQSCHRSLPVF 226
QY 223 --KLYHS-----DACSRVAVSLRCI-----ACGVNL 248
Db 237 CHSFFYDTPDLGELDIYAKSHACQKLTCTNVRQFFLYTPAASCNBGRKCYLKL 296
QY 249 NSNR-----QSRVIGESALPGAMPVQSLAVQNV 278
Db 297 SSNSGSPKILHGRGIGSYTLRLCKMNECTTKIKPRIVGTASVREMPQVTLHTTSP 356
QY 279 ---HVCGSITPEMAYTAHCYKPLNPNMWTAFAGTLRSPM-----FYGAGVYER 330
Db 357 TORLCCGSIIGNQWILTAHCF-YGVESPIILRVYGLINQSEIKEDTSFEG---VOE 411
QY 331 VISHPNDSKTKNNDIALMLOKPLTFNDLVKPYCLPNDGMLOPEOLCWSGATEEK 390
Db 412 ITHDQYKMAESGDYDILKLTETVNTDSRPLCLPSKODRNYITDCVYTGKGYKLR 471
QY 391 GKTSEVINAARVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSCQDGSGLPVTSKN 450
Db 472 DKIONTLQAKRIPLVNNECQKRY-RGHKITHKMICAGYREGKDAKDGSGPLSCKHN 530
QY 451 NIMMLIDGTSWGSACAKAYRPGVGNVVEFTDWIYRQMR 490
Db 531 EVMHLVGTITSWGEGCAQREPRGYTVNVEYVDWILEKTA 570

```

```

RESULT 13
Q9Y495 PRELIMINARY: PRT: 310 AA.
AC Q9Y495;
DT 01-MAY-2000 (TREMBLREL. 13, Created)
DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
DE DIGITAL INTESTINAL SERINE PROTEASE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,
RA Walters J.R.;
RT "Characterization of a novel murine intestinal serine protease,
RT DSP."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ243866; CAB56465.1; -
DR HSSP: P20231; IAAO.
DR INTERPRO: IPR001254; -
DR INTERPRO: IPR001314; -
DR PRAM: PF000089; trypsin. 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR KMW Protease.
SQ SEQUENCE 310 AA; 33701 MW; F828BC7F6D25303F CRC64;

```

```

Query Match 19.5%; Score 529; DB 11; Length 310;
Best Local Similarity 41.7%; Pred. No. 4.8e-40;
Matches 105; Conservative 42; Mismatches 91; Indels 14; Gaps 6;
QY 244 CGVNLNRSRQSRIYVGSALPGAMPVQSLHV-ONVHVGCGSITPEMAYTAHCVCKPL 302
Db 28 CG---HSRDAKIVGGQDALEGPMPVQSLMTITGDCIGCGSLHETVLTAAHCFRSL 84
QY 303 NNPMHTAFAGILRQSMFYAGYQVKEVISHPNY-DSKTKNNDIALMLOKPLTFNDLV 361
Db 85 NPSYHYHVKVGLTSLLEPHSTLVAVRNIFVHPYLTAAADSSGDIALVQDTPLRPSQFT 144
QY 362 KPVCLPNDGMLOPEOLCWSGATEEKGKTSSEVYLNAAVLLIETQRCNSRY----- 414
|||||

```

Db 517 TPQTNPRAGLEKNYCRNPDGVDVNGPMCYTMNPRKLY--DYCNIPLCASLSSFECKPOVE 574
QY 251 SRO--SRIVGESALPGAMPQOYSLHVQ--NVHVCGSITTPEMIVTAHCVKPLNPNW 306
Db 575 PKCQGRVYGGCVANPHSNWQISLRTFSGOHFCGTLISPEWLTAAHCLEKS-SRPE 633
QY 307 HMTAFAGILRQSFMYGAGYQ---VEKVISHPNYDSKTKNNDIALMKLQPLTFNDLVKP 363
Db 634 FYVYILGAHEERIL--GSDVQOIAVTKLVLEPN-----DADIALKLSRPATITDNIVIP 685
QY 364 VCLPNPGMMLQPEQLCHISGNGATEEKGT-SEVLNAKAVLLIETORCNSRYVDNLITP 422
Db 686 ACLPSPNYVADRTLCTYITWG--ETKGTGAGRLKEAOLPIENKVCNRAEYLNNRVKS 743
QY 423 AMICAGFLOGNDSCGDSGGLYTSKNINIMWLIIGTSMGSCAKAYRPGVYGNVWVETD 482
Db 744 TELCAGHLAAGIDSCGDSGGLVCFEKDKYILOGVTSWGLGCAHPNKPQYVYVRSRYVN 803
QY 483 WIYROMRAD 491
Db 804 WIEREMRND 812

Search completed: May 8, 2001, 11:57:28
Job time: 143 sec